THE
QUARTERLY JOURNAL OF ECONOMICS

FOUNDED 1886

The Quarterly Journal of Economics

Published by Harvard University

Books, periodicals, and manuscripts to be addressed, THE QUARTERLY JOURNAL OF ECONOMICS, M-12, Littauer Center, Cambridge, Mass.

Business letters to be addressed, HARVARD UNIVERSITY PRESS, 44 Francis Avenue, Cambridge, Mass.

Subscriptions $5.00 a year. Three years, $14.00. Single copies (except those published in two parts) $1.50, postpaid. A special subscription rate of $3.00 is offered to students. Such special subscriptions may be for one, two, three, or four years. Those desiring to take advantage of this reduced rate should include with their remittance a statement from an officer of their college or university that they are *at that time* "regularly enrolled students who have not yet received the Doctor's degree."

CONTENTS FOR NOVEMBER, 1954

THE
QUARTERLY JOURNAL
OF ECONOMICS

SCHUMPETER'S HISTORY OF ECONOMIC ANALYSIS*

By LIONEL ROBBINS

I

The contents of this large and (physically) very heavy volume are correctly indicated by its title. It is not a history of theories of policy; nor is it a history of economic systems; these matters are discussed only in so far as they throw light on the central subject — the evolution of those techniques of thought which enable us to describe and explain the economic aspects of reality. The plan falls into five parts. Part I (pp. 3–50) is devoted to a vindication of the autonomy of the subject and a discussion of its relations with other branches of knowledge. Part II (pp. 51–378) traces developments from the beginnings in Greco-Roman times to the promulgation in the last half of the eighteenth century of the Physiocratic and Smithian systems. Part III (pp. 379–752) deals with the period 1780 to 1870, so covering the English Classical school in its full development. Part IV (pp. 753–1138) takes us down to 1914 "and later," and ranges far and wide over the evolution of analysis since the Jevonian revolution. Part V (pp. 1139–84) is a fragment and deals in a sketchy way with certain very recent developments.

The genesis of this most ambitious enterprise lies a long way back. The monograph, *Epochen der Methoden- und Dogmengeschichte*, which as a young man Schumpeter contributed to the Tübingen *Grundriss der Sozialökonomik*, had at once taken its place as incontestably the best short introduction to its subject; and several attempts

had been made to induce its author to consent to an English translation. But he always held these at arm's length, declaring himself to be desirous of making some revision before allowing such a version to appear. Eventually, in 1941, when he had finished his *Capitalism, Socialism and Democracy*, he embarked upon this task, apparently conceiving it as a convenient side line to his main concern of preparing a major work on theory. Gradually, however, his interest developed; and thus for the last nine years of his life, so Mrs. Schumpeter tells us, it became a dominating preoccupation.

At the time of his death, his labors had resulted in the work which is embodied in the present volume. But the materials were far from their present form: they were scattered in many manuscripts in many places, some typed and corrected, some only typed, some merely in longhand. We owe it to Mrs. Schumpeter, devoted to this task through a prolonged illness even to the last weeks of her life, that the manuscripts have been assembled in their present form — a coherent, if not finally completed, book. The story of this reconstruction is itself almost a minor epic. The annals of our profession abound with the examples of wives whose unselfishness has made possible their husbands' books; and more than one wife has successfully supervised the publication of posthumous works. But it is doubtful whether there is any parallel to such a labor as this — a shining example of what comradely devotion can achieve. It is quite true that the book is not finished in the sense in which the author would have finished it had he lived to see it through the press. A few chapters are unwritten; several are unfinished. Occasionally even sentences tail off without a conclusion — the editor conceived her "editorial task to be the simple one of presenting as complete and accurate a version of what J.A.S. actually wrote as possible but not attempting to complete what he had not written." But, as with the Leonardo *Adoration* in the Uffizi, the grand proportions are there: and if some details are unfinished and others lacking, if here the perspective is only indicated by a sketch and there alternative versions blur the outlines, yet it is essentially an integral creation, to be judged and enjoyed as a whole.

II

There were many reasons why Schumpeter should have been peculiarly well fitted for an enterprise of this sort. He was born and brought up in one of the most brilliant and cultivated societies of Europe before the decline, at that time one of the chief centres of speculation in theoretical economics. As a young man he had travelled widely and studied in many other places; he had personal con-

tact with many of the founders of modern economic analysis elsewhere. He had thus exceptional opportunities for appreciating the unity in apparent diversity of different traditions and different schools of thought. His experience, moreover, was not wholly academic. He had some knowledge of the world of affairs, both of politics and of business, although, if common report is to be believed, his achievements in these fields were not of any conspicuous excellence. He thus approached his subject with that modicum of practical experience without which, although it is no integral part of the subject, so much economic analysis is apt to be sterile. Then, in his middle age, when still at the height of his powers, a great American university had called him to a position, remote from these temptations, with exactly the right combination of academic duty and leisure, in which, all passion spent — or nearly spent — these powers could express themselves without distraction.

There were gifts, too, of intellect and temperament reinforcing these environmental advantages. He had a first-rate mind. He had himself made contributions of distinction to the corpus of modern analysis. Unlike many who have written on the history of this subject, he knew what it was about from the inside; he spoke with the authority of a high practitioner. He was, moreover, an excellent expositor and showman; in our profession, with the single exception of Keynes, he was probably the best talker of his generation. Then, although he was not incapable of the slightly feline dig, he was essentially fair and good-tempered. Himself the product of many streams of influence, his whole intellectual bent was eclectic — an excellent thing in an historian of thought. Indeed, it is arguable that at times this quality was present almost to excess; there are places in this book where the distribution of praise is on so lavish a scale that it is hard to suppress a feeling of being present at some very high-grade get-together at which the exceptionally knowledgeable chairman is almost too meticulously careful not to leave unmentioned any guest, however humble. Finally, with all his erudition, he was no pedant. He felt none of that compulsion often felt by men of high ability to see good only in one line, no matter what else is excluded. He knew enough of the world to know why some men of sense are irritated by economics; but he knew so much of economics that he knew too at what stage this irritation ceased to be justified. All of which, it must be agreed, constituted quite exceptional qualifications for writing the history of analysis.

Let it be said at once that the result, slightly unfinished as it is, is something which is without serious rival. The history of economic analysis has attracted some very first class talent: some of the con-

tributions — such works, for instance, as Professor Viner's *Studies* or Mr. Sraffa's *Ricardo* — are outstanding in the scholarship of historical studies in general. But for the most part the work of high excellence here has taken the form of special studies rather than general history. There is very little that is absolutely first class in this sphere and nothing that will stand comparison, either in respect of scale or insight, with this work of Schumpeter's. This is a really serious and mature treatment of the subject. And, although every expert reader will probably find something to grumble about in the parts which deal with the periods or authors he knows best, yet taking things in the large, it is safe to say that it will be a long time before it is displaced as the leading work in its field. Whether at the same time it is a good introduction is a matter of greater dubiety. If one has some familiarity with the material discussed, it is not easy to decide whether all the copious references and allusions are necessarily well suited to convey a clear picture to the beginner. My own guess would be that very frequently they are not. But this is no criticism of the book on the plane on which it elects to move. The right way to begin to acquire a knowledge of the subject is to read the original texts themselves. Reading about them, pondering comments on their significance, comes later; and it is at this stage that Schumpeter comes into his own. And if any beginner, not knowing his way about, is in need of a compendious guide, he can always read the *Epochen* — now at last fortunately available in translation.

If this verdict is correct, if here at last we have something which can be regarded as a standard work on its subject, it follows that criticism must be concerned chiefly with matters of particular perspective and detail. That, indeed, in the main is the substance of what follows. But before thus closing up to the canvas, it is perhaps worth asking whether there are no critical comments to be made· of its general appearance and execution.

In my judgment there are two — not, I think, of such importance as seriously to detract from the total achievement, but still perhaps deserving to be put on record.

First, I must confess that I find it hard to resist a certain impression of undue length. By this, I hasten to say, I mean no derogation of its aim or its merit as a treatise on the grand scale — despite the canons of a somewhat short-winded age, there *is* a certain virtue in bigness, provided that it is proportionate to its object. I refer rather to a certain tendency to discursiveness, a certain lack of control of the impulse to elaborate, which at times at least seems to be excessive. I suspect that this book could have been a couple of hundred pages shorter with advantage. This is perhaps partly due to the unfinished

state of the manuscript: a final revision would almost certainly have involved a slimming process here and there — it would be unfair not to take this into account. But partly it is probably due to a more fundamental cause; I suspect that Schumpeter's literary habits did tend to over-elaboration — think, for instance, of the *Wesen und Hauptinhalt* — or even, dare I say it, of *Business Cycles*. I ought perhaps to add that I do not think that the result is boring; the many long summer days that I have devoted to reading these 1184 pages have been most agreeable. It was always pleasant to hear Schumpeter hold forth, even if now and then his points were a little otiose. But prolixity is a fault, even though it may sometimes be a fault on the good side; and I think that this book is a little prolix.

Secondly, I suspect that now and then there is a tendency to overemphasis and even to paradox in connection with certain reputations. This is not an easy point to make, since, taking the book as a whole, one of its chief merits is its excellent perspective. Moreover, it may well be that some differences with Schumpeter here are irreducible differences of temperament and local affiliation. Nevertheless, I do think that from time to time there are errors of this sort, so that, so to speak, certain figures suddenly pop out of the canvas and lose relation with their surroundings. Thus, for instance, on page 465 in the course of a well-merited eulogy of von Thünen we are surprised to read that "If we judge both men exclusively *by the amount of ability of the purely theoretical kind* that went into their work, then, I think Thünen should be placed above Ricardo or indeed above any economist of the period with the possible exception of Cournot." Now, of course, arguing about class lists at this level is not a particularly profitable occupation: and arguing about von Thünen in particular is highly esoteric. But surely there is something wrong about a judgment which, leaving aside von Thünen, even on Schumpeter's somewhat specialized criterion, would place Cournot above Ricardo. Cournot was certainly a highly talented man whose formulations of the demand function and the theory of monopoly were contributions of absolutely first class importance. But let us apply the good Austrian test of the so-called loss principle. Suppose the theoretical contributions of Ricardo and Cournot to be alternately withdrawn — which would involve the greater impoverishment? Can there really be any hesitation about the answer? Certainly the relative amounts of space given to the work of these two authors by Schumpeter himself would seem to leave no doubt about it. Similarly, later on in the book, we find pre-eminence in his period assigned to Walras rather than to Marshall[1] — although it must be said at once that Schumpeter

1. *Op. cit.*, p. 827.

has many fine things to say about the latter. Again, I cannot but feel a certain artificiality of emphasis. Undoubtedly Walras was a great economist, perhaps among the greatest. But is it really to be argued that the loss of his contribution, outstanding though it was, could outweigh the loss of Marshall's? We may agree that Walras had a raw deal during his life and is probably not sufficiently appreciated even in our own day. But I submit it is to get our perspective wrong if, in the laudable effort to rehabilitate him, we rate him as high as all that. But more of this later.

III

Part II, where the history really begins, runs from Plato and Aristotle to the physiocrats and Adam Smith. It is easy to see the advantages of this arrangement. The vision of streams of fragmentary pieces of economic analysis springing from the treatises of moral philosophers and the *ad hoc* utterances of administrators and pamphleteers and culminating in the eighteenth century in the discovery of the system in economic life, is a fine one and opens up grand perspectives. But to exhibit it without a break in time has two disadvantages: it tends to minimize the degree of achievement in the eighteenth century discovery of system: Quesnay and Adam Smith (and Cantillon before them) were not just two more moral philosophers or pamphleteers. And to have a break after Adam Smith but not before tends to suggest a certain divorce between the *Wealth of Nations* and the Classical writings of the nineteenth century which may foster misunderstanding; the unity of the English Classical school, either as regards analysis or as regards policy, is not to be understood save in the light of a common acceptance of the general outlook of Adam Smith. In the *Epochen* there were two parts covering this period: Part I, *Die Entwicklung der Socialökonomik zur Wissenschaft* which dealt with the philosophers and pamphleteers, and Part II, *Die Entdeckung des Wirtschaftlichen Kreislaufs*, which, as its name implies, dealt with the discovery of the economic system. I am inclined to think that for the main purpose of the book, although perhaps not for the treatment of such subjects as money and interest, this arrangement was superior.

So far as the philosophers and pamphleteers are concerned, the treatment seems to me to be in the main well proportioned and just. I say this with some reserve, since so far as the scholastic writings are concerned, I have never lived with the original texts long enough to form more than the broadest impressions; and there are a great many aspects of the mercantilist literature about which I still find it very

hard to make up my mind. It is quite clear that the rejection of most of this literature as worthless was a gross error: if the classical economists have suffered from overcrude interpretations of their point of view, it must be admitted that they themselves were guilty of the same kind of unfairness in regard to their predecessors. But I am fairly sure that more recent interpretations, either by members of the Historical school or by Keynes, tend to err in the opposite direction: the theory of the barrenness of money did not rest upon a liquidity preference theory of interest, nor did emphasis, in a régime of advancing prices, upon the accumulation of treasure necessarily rest upon dread of deflation. In spite of a certain love of paradox and of the *recherché*, Schumpeter treads very delicately here. I suspect that some of those who know best these branches of the literature would tend to the view that his interpretations were still too charitable; and there is doubtless much scope for dispute about particular interpretations. But in general his treatment must command respect: if it is not a final verdict, it provides at least agenda for grown-up talk.

On the eighteenth century developments I find him less impressive. He has none of Hume's feelings about the physiocrats — very much the contrary indeed;[2] and he provides an account of Quesnay's position which I venture to say is much more impressive and very much more intelligible than anything that that very gnomic expositor ever wrote himself. But he still leaves one reader at least with the feeling that, individually and as a group, they have sometimes been overpraised and are indeed overpraised by him. Doubtless the *Tableau Economique* can be construed to stand for an idea of great power and importance — though I fancy that there is more jobbing backward in this interpretation than we are often prepared to admit. But it is surely a significant circumstance that among the rest of the group the scientific aspect of the *Tableau* had hardly any influence; it was an object of worship, the symbol of a cult rather than a stimulus to further development. If profundity of thought and scientific detachment are the qualities to be valued, the work of Petty and of Cantillon seems to me to be on a much higher level.

As for the treatment of Adam Smith, I am inclined to regard it as one of the few real lapses in the book. There can be little doubt that Adam Smith was one of Schumpeter's blind spots. In an earlier, though not so much earlier, essay, he describes Smith's achievement as a " 'burst into publicity' — the *Wealth of Nations* was not more

2. "They are, indeed, the set of men the most chimerical and arrogant that now exist, since the annihilation of the Sorbonne . . . I wonder what could engage our friend, M. Turgot, to herd among them." *Letters of David Hume*, ed. J. Y. T. Greig, II, 205.

than this.''[3]   And in the present volume he actually says that nobody who had not read them would credit the author of the *Wealth of Nations* with the power to write the posthumously published essays.[4] It would be possible, of course, to quote sentences which in themselves convey high praise.   But I do not think that it is open to question that the general impression which is left by the treatment of Adam Smith is one of slightly patronizing, slightly derogatory debunking — the man was a good systematizer of other people's ideas, an excellent expositor, lucky in his period, whose "very limitations made for success."[5]   Why this was so, whether it was because of Smith's coolness about Political Arithmetic —he "took the safe side that was so congenial to him" — [6] or whether it was because of some chemical disaffinity of temperament —— " a fact which I cannot help considering relevant, not for his mere economics, of course, but all the more for his understanding of human nature — that no woman, excepting his mother, ever played a role in his existence,"[7] — it is difficult to say. But the existence of some nonrational inhibition is incontestable.   It is not the first time that a scholar with Schumpeter's background has been baffled and irritated by the unpretentious clarity and restraint of eighteenth century English moral philosophy.

IV

Part III, which occupies nearly four hundred pages, contains the fullest and most systematic account to be found anywhere of the Classical system in its full development.   For this much gratitude is due.   But I have some reservations about matters which are not altogether unimportant

First a matter of background influence—the role of Utilitarian moral philosophy.   I am inclined to think that Schumpeter got this all wrong.   I am not thinking here of his description of this outlook as "the shallowest of all conceivable philosophies of life"[8]; that can conceivably be overlooked as a lapse into Neo-Hegelian cliché which might perhaps have been expunged in a final revision.   I am thinking rather of his suggestion that the Utilitarian background was to be conceived as confined to Bentham and the two Mills.[9]

3. ''Professor Taussig on Wages and Capital,'' *Explorations in Economics*, p. 213.

4. *Op. cit.*, p. 182.

5. *Ibid.*, p. 185.

6. *Ibid.*, p. 212.

7. *Ibid.*, p. 182.

8. *Ibid.*, pp. 133 and 407.

9. *Ibid.*, p. 408.

This seems to me very questionable. I would have said rather that the background of the entire school, from its beginnings in Hume's *Essays* right through to Cairnes and Sidgwick, was through and through Utilitarian—and this stands even for Adam Smith whose explicit moral philosophy had a somewhat different complexion. In my judgment it is a mistake to regard Bentham and the Mills as *the* characteristic English Utilitarians. They represent—or at least Bentham and James Mill represent — a peculiarly severe and idiosyncratic version of a general outlook which, from a philosophical point of view, has had much more persuasive advocates both before and since.[1] And the habit of judging actions and policies by their consequences rather than by reference to some intuitive norm, which is the salient characteristic of this outlook in contrast to that of the continental metaphysicals, was common to almost all the English speaking economists of the age. The contrast is well exemplified in Nassau Senior's conversation with the de Broglies:

"This led us to talk of Bowood and from Bowood we went to Bentham. The Duke had never seen his Bowood correspondence: indeed he seemed to be little acquainted with his works. The Prince knew more of them but admitted that he had not studied them accurately.

Prince de Broglie — There are valuable hints in them as there must be in the works of a man of diligence and originality, but they scarcely repay the labour of mastering a system based on error.

Senior — Do you reject then utility as the foundation of morals?

Prince de Broglie — Certainly I do. It is generally rejected in France.

Senior — And what do you substitute?

Prince de Broglie — Our innate feelings of right and wrong."[2]

It might be argued, of course, that Senior was just the neutral interrogator, as he so frequently was in his conversations. But I am pretty sure that in this instance this interpretation would be wrong. The question, what is to be done if we reject considerations of utility, is typical of the attitude of the main tradition of English political economy. I hasten to add that I fully agree with Schumpeter that the logical coherence of their analytical propositions does not stand or fall with this background.[3]

1. Schumpeter's footnote (p. 408) in which he tries to rescue Sidgwick from this classification is not one of his happier efforts.
2. Nassau Senior, *Conversations with M. Thiers, M. Guizot and other Distinguished Persons*, II, 176.
3. Painful experience compels me to register the hope that the above piece of historical diagnosis will not be taken as a manifesto of personal affiliation.

Next comes a matter of position and influence. It is Schumpeter's contention — or at least, so it appears at a certain stage of his argument — that Ricardo's influence has been overstressed, that there were no conspicuous Ricardians save James Mill, McCulloch, and De Quincey, and that, by the beginning of the thirties, the school had spent itself and was no longer a living force. There were the Ricardian socialists, of course, there were Marx and Rodbertus — and Rossi! But the main stream of English Classical analysis had emancipated itself from this influence. This is a point of view which recently has had some support in other quarters. It is, therefore, perhaps deserving of more thorough examination.

A great deal of the paradox in this position begins to vanish if we are prepared sufficiently to narrow our concept of what the Ricardian position was. Of course, if we interpret this position as chiefly involving the assumption of an absolute standard of value, on the lines attacked by Bailey[4], and the peculiar doctrine regarding the relation between "wages" and "profits" (i.e., proportionate wages and proportionate profits) which from that day to this have been such a stumbling block to correct interpretation, it is perhaps just to say that it died with James Mill — with De Quincey's *Logic* as a very belated posthumous progeny — and we can even bring ourselves to stomach the astonishing proposition that J. S. Mill was not really Ricardian.[5]

But this surely is an intolerable restriction. A juster definition, I suggest, would leave the influence of Ricardo over a very wide field quite dominant throughout the whole of the Classical period. It was a stroke of expository genius, on Schumpeter's part, to make Senior's fundamental propositions the peg on which to hang his central analysis of the Classical position. But how much of Senior's work would have seen the light in the form in which it did, were it not for the influence and example of Ricardo? It is true that at many points Senior is critical of Ricardo. But where else in his book does Schumpeter restrict the significance of the term "influence" to slavish reproduction? What is relevant in this connection is not whether the analysis of value and distribution of Senior and J. S. Mill — not to mention money and international trade — followed Ricardo in all respects but

Personally, I am content to be something of a pluralist in moral philosophy. I think a utilitarian calculus is quite a good first approximation to many of the workaday problems of social policy. For the ultimate problems of personal conduct I am much less disposed to regard it as helpful.

4. It is more than doubtful whether Ricardo would have pleaded guilty to this indictment.

5. *Op. cit.*, p. 529.

whether it conformed to the type of analysis of which his works are the archetype. I cannot help thinking that, once he had released his paradox, Schumpeter would not really have disputed this position. At any rate by page 677 we find him arguing, apropos of the theory of rent, that "so great was Ricardo's success that even some writers who adopted Say's scheme in other respects, inserted into it a Ricardian treatment of rent without betraying any symptom of logical discomfort."[6]

In this connection I am inclined to suggest that much too much can be made of some of the things said at the celebrated evenings of January 13th and April 15th, 1831, at the Political Economy Club at which the question was debated "What improvements have been effected in the Science of Political Economy since the publication of Mr. Ricardo's great work: and are any of the principles first advanced in that work now acknowledged to be correct". A recent writer[7] in *Economica* has made the account of the proceedings given in Mallet's diary a reinforcement for a diagnosis of rapid decline of Ricardian Economics. In fact, however, Mallet's account is by no means decisive on this point. The general conclusion, at the end of the second meeting, seems to have been that "Ricardo is a bad and obscure writer, using the same terms in different senses: but that his principles are in the main right."[8] What is perhaps, however, even more significant and what has apparently escaped Dr. Meek's attention is that Torrens, who on this occasion was the ringleader of the anti-Ricardians, subsequently repented of this attitude and published one of the most candid retractations in the history of economic discussion. I do not think that this covered his critique of the Ricardian search for a measure of value: he continued to give praise to Bailey. But it certainly covered the rest. The Introduction to the *Budget* is largely an attempted vindication of Ricardo on profits and a disavowal of earlier strictures on this doctrine and on the theory of rent. "Some of the commentators on the doctrines of Ricardo appear to have fallen into the misconception that, in altering his nomenclature, and in modifying his principles as varying circumstances required, they refuted his theory of profits. In this censure I include myself."[9]

I now come to a number of more detailed points. In his discussion of population theory, Schumpeter gives the impression that the law of diminishing returns had nothing to do with the original

6. The writers were Roscher and J. S. Mill.
7. R. L. Meek, "The Decline of Ricardian Economics in England," *Economics*, Feb. 1950.
8. Centenary Volume of the Political Economy Club, p. 225.
9. *The Budget*, p. xxxvi.

formulation of Malthusian doctrine. This view, of course, has very formidable supporters. No less an authority than Edwin Cannan, who was so seldom wrong on his texts, gave it a classic form in his dictum that "to imagine that the *Essay on the Principle of Population* was ever based on the law of diminishing returns is to confuse Malthusianism as expounded by J. S. Mill with Malthusianism as expounded by Malthus."[1] Nevertheless, I do not think that he was right. Admittedly, the mere comparison of geometrical and arithmetical potentialities of increase does not itself necessarily involve this relationship. But the idea underlying the comparison is of the same order of conceptions; and it is surely no mere coincidence that in another connection Malthus was one of the first to formulate the so-called law. This, of course, has been perceived before and, indeed, is clearly the basis of the view which is attacked by Cannan; up to this point, perhaps, the issue might be regarded as undecided. But the textual evidence seems to indicate otherwise. It is true that there is no mention of diminishing returns in the *first* edition of the *Essay*. I am not so clear, however, that this can be asserted of the second edition, where, on page 7, we are told that "in proportion as cultivation extended, the addition that could yearly be made to the former average produce must be gradually and regularly diminishing." What, however, seems to me decisive is the language of the *Summary View of the Principle of Population*, the extract from his article in the Supplement to the *Encyclopaedia Britannica* which Malthus issued in 1830.

"It has already been stated, that while land of good quality is in great abundance, the rate at which food might be made to increase would far exceed what is necessary to keep pace with the most rapid increase of population which the laws of nature in relation to human kind permit. But if society were so constituted as to give the fullest scope possible to the progress of cultivation and population, all such lands, and all lands of moderate quality, would soon be occupied; and when the future increase of the supply of food came to depend upon the taking of very poor land into cultivation, and the gradual and laborious improvement of the land already cultivated, the rate of the increase of food would certainly have a greater resemblance to a decreasing geometrical ratio than to an increasing one. The yearly movement of food would, at any rate, have a constant tendency to diminish, and the amount of the increase of each succeeding ten years would probably be less than that of the preceding."[2]

1. *Theories of Production and Distribution*, p. 144. In my salad days, in a *Festschrift* for Cannan, I echoed this view very strongly. The curious may consult *London Essays in Economics*, pp. 104–5.

2. *Op. cit.* Reprinted in Glass, *Introduction to Malthus*, p. 139.

Clearly this is not a very strict formulation. But at least it makes it very difficult to maintain that there was no connection in Malthus' mind between his ratios and diminishing returns to labor applied to land. The good sense of the matter seems to be admirably expressed by Dr. McCleary when he says that "although Malthus did not explicitly base his arithmetical ratio on the law of Diminishing Returns, the concept of diminishing returns was present and influential in his mind."[3]

I find myself equally unable to agree with Schumpeter upon an important aspect of the Ricardian theory of value. On pages 600 and 601 he has a long disquisition on Ricardo's perversity in opposing the supply and demand theory for the explanation of long-run values, contending that he was "completely blind to the nature and the logical place in economic theory of the supply and demand apparatus . . . had he but stopped to ask *why* exchange values of commodities would be proportional to the quantities of standard labour embodied in them, he would, in answering this question, have found himself using the supply and demand apparatus by which alone (under appropriate assumptions) that law of value can be established."

Now there can be no doubt that Ricardo had himself to thank for a great deal of the misunderstanding which has arisen in this connection. To have written a whole chapter *against* the supply and demand theory, without defining more clearly exactly wherein lay the nature of his opposition, was simply asking for trouble. Moreover, it is clear that once we get away from the very peculiar basis of the Ricardian theory, there is a wide sense in which the cost of production which may be opposed to supply and demand as an explanation of normal value, is itself determined by the forces underlying supply and demand for factors and their products; the special theory which sidetracks all this proves to be based on intolerably restrictive assumptions. Nevertheless, the idea that while he accepted supply and demand as determining market values, he ignored their operation in the determination of natural values is a complete mare's nest, despite the very high authority, from Senior onwards, by which it has been supported. This is conclusively shown by the very *Notes on Malthus* to which Schumpeter refers us; the relevant note stands out half a mile:

"Mr. Malthus mistakes the question — I do not say that the value of a commodity will always conform to its natural value without

3. *The Malthusian Population Theory*, p. 111. This excellent essay may be strongly recommended as the sanest study of this highly controversial subject. It is worth observing that Dr. McCleary's view has the support of the high authority of Marshall.

an additional supply, but I say that the cost of production regulates the supply and therefore regulates the price."[4]

V

Part IV, 1870–1914 (and later) is in effect a review of modern economic theory since Jevons and, in my judgment, is the best thing in the whole book. These were the developments, which were most congenial to Schumpeter's own temperament, and it was here that his roving eclecticism and the conciliatory side of his character showed themselves to their best advantage. The story is told both in terms of persons and of theorems. The leading personalities are reviewed; some of the sketches here are astonishingly felicitous: for instance, I do not think that Bohm-Bawerk and his contribution have ever been more persuasively depicted. The general character of theoretical developments in different fields is set forth. A special chapter is devoted to elucidating the contents of equilibrium analysis with the *Elements* of Walras playing the part here which, in the section on the Classical theorems, was allotted to Senior's four propositions. There are special studies of the evolution of the idea of the production function and the vicissitudes of the theory of utility, and a separate chapter on money and cycle theory. If any enquirer wishes for a unified account of what modern theory (not merely to 1914 but, as the title says in brackets, "and later") has been about, it is here for him to read.

Perhaps the most admirable feature of all this is its cumulative demonstration of the underlying unity, amid apparent diversity, of all this speculative activity. Schumpeter excelled at this. It was his real conviction, as I think he once said, that there are fundamentally no schools nowadays, but only good economists and bad economists, and he brings this out with extraordinary sympathy and force in regard to the work of this period. Perhaps from time to time this is overdone. There was, for instance, more at issue in the controversy about real costs than Schumpeter is willing to concede — though I should be perfectly prepared to agree, and indeed to argue, that it can chiefly be resolved into a question of appropriate approximation. Again the undeniable hostility of the Lausanne school to any form of partial equilibrium analysis, however safeguarded and qualified, was a very real barrier between members of that group and their colleagues elsewhere; and their intransigence in that respect, I should say, quite as much as the blameworthy indifference of the outside world, was

4. *Notes on Malthus*, Sraffa edition, pp. 48–49.

one of the chief reasons why the good side of their work had to wait so long for recognition. But in the main the assertion of a common subject matter and common methods is sound and it is no small service thus so definitely to have distinguished the wood from the trees.

In this general perspective perhaps the least satisfactory feature is the treatment of Marshall. This is a delicate and difficult matter to elucidate: delicate, since I am fully aware that the main criticisms in this notice may be represented as defenses of a local tradition, difficult, since in many respects Schumpeter writes so well about Marshall — how refreshing, for instance, it is to see it clearly recognized that markets peculiar to individual firms are one of the important features of Marshall's analysis.[5] Nevertheless there are misconceptions and criticisms on which some comment seems to be called for.

Perhaps the best method of approaching the core of discomfort here is by way of examining Schumpeter's contention that Marshall was ungenerous, that he did not pay sufficient tribute to the originators of some of his most fundamental ideas.[6] It seems to me that there is very little in this. I should not be prepared to contend that Marshall was always an altogether agreeable character; he was obviously a bit of a Turk at home and his behavior to Sidgwick on one occasion certainly calls for explanation.[7] But there is really no evidence at all that he did not handsomely acknowledge the debts of which he was conscious. His references to Cournot and von Thünen probably did more than anything else to revive and spread the reputation of these authors and his solicitude for the position of Ricardo even comes under rebuke from Schumpeter. It is true that he wrote a severe review of Jevons. But there is no reason to doubt his own explanation that he was angered by the treatment of Ricardo and disappointed by the fragmentary nature of the *Theory*. Elsewhere he paid ample tribute to the lively and powerful intelligence of its author.

In this connection it is perhaps worth quoting from a manuscript note which is reproduced in the *Memorials;* it gives a more vivid

5. This could perhaps with advantage have been reinforced by explicit quotation of the paragraphs (8th ed., pp. 841–50) in the mathematical appendix where the distinction between marginal revenue and price is very clearly set out, together with reasons for neglecting it in analysis of the more competitive part of the field. And if, at this juncture, our author had been in a more hortatory mood, he might have pointed out that, besides being set out very clearly in Cournot, the description of monopoly equilibrium in terms of equality between marginal revenue and marginal cost was even used in such a popular textbook as Chapman's *Outlines* (p. 177).

6. *Op. cit.*, pp. 838 *seq.*

7. See *Henry Sidgwick: A Memoir,* by A. S. and E. M. S., pp. 394–96.

account of a state of mind than any amount of secondhand conjecture. "I looked," wrote Marshall, "with great excitement for Jevons Theory, but he gave me no help in my difficulties and I was vexed. I have since learned to estimate him better. His many-sidedness, his power of combining statistical with mathematical investigations, his ever fresh honest sparkling individuality and suggestiveness impressed me gradually and I reverence him now as among the greatest of economists. But even now I think that the central argument of his theory stands on a lower plane than the work of Cournot and von Thünen . . . . My youthful loyalty to him [Ricardo] boiled over when I read Jevons' *Theory* . . . . I have a vivid memory of the angry phrases which would force themselves into my draft, only to be cut out and then reappear in another form a little later on, and then to be cut out again."[8] This is not the utterance of an ungenerous spirit.

A similar quotation may serve to indicate his position regarding Dupuit, another victim of neglect, according to Schumpeter. "The notion of an exact measurement of Consumer's Rent was published by Dupuit in 1844. But his work was forgotten; and the first to publish a clear analysis of the relation of total to marginal (or final) utility was Jevons in 1871, when he had not read Dupuit. The notion of Consumer's Rent was suggested to the present writer by a study of the mathematical aspects of demand and utility under the influence of Cournot, von Thünen and Bentham."[9] This surely is complete enough recognition. Schumpeter speaks of "footnotes in the wrong place."[1] But this footnote is just where it should be — after the exposition of the doctrine of Consumer's Surplus in the text. He might have asked with greater cogency why it was omitted in later editions. But the charitable solution here would rest upon Marshall's conceptions of the exigencies of a textbook in popular use; and when one thinks of the way in which Marshall consistently underplayed his own originality, seeking to present the most pathbreaking constructions as notions assumed implicitly by the masters of the past, it is reasonable to assume that he did not attach overwhelming importance to the acknowledgment on all possible occasions of all possible patents.

There remains the question of his attitude to Walras, to whom reference in the text of the *Principles* is very sparse and that not at all on the main Walrasian contributions. Here we reach considera-

8. *Op. cit.*, p. 99.
9. *Principles*, 2d ed., p. 184.
1. *Op. cit.*, p. 840.

tions much more important than scientific etiquette and the presence or absence of a magnanimous disposition. We have to consider the nature of the respective contributions of the men concerned and Marshall's probable attitude towards that of Walras.

Now there can be no doubt whatever, as Schumpeter himself acknowledges, that Marshall was fully seized of the conceptions of general equilibrium. The attempt, sometimes made by members of the Lausanne school, to oppose a supposititious Marshallian partial equilibrium system to a Walras-cum-Pareto general equilibrium system just will not bear examination. The opening chapters of Book VI (Preliminary Survey of Distribution) together with the famous Note XIV in the Mathematical Appendix are essentially the presentation of an approach as ambitious in its scope and vision as anything presented by Walras, although much more concise in expression. The idea that there can have been anything particularly novel to Marshall in the Walrasian equations is not at all plausible. He certainly knew about utility theory and, as we have seen, he had acknowledged his sources — which were not Walrasian. The influence of demand on the pricing of productive services, explained in terms of the net product and marginal product principles, was a central feature of his own system. Explanation of interest, and hence capitalization, in terms of the productiveness of permanent capital instruments and the discount on the future restraining their supply, was equally prominent in his thought; and the filling out of this statical theory by a cash balance approach to the theory of money had figured prominently in his evidence to the Gold and Silver Commission — some time before Walras had fitted it into his own system. Charged with neglect of the Walrasian contributions, I am sure that Marshall would have defended himself in this manner.

But if asked why nevertheless he did not make more specific reference to their treatment by Walras, I am inclined to think that he would have moved to the offensive. He did not make more specific reference to the Walrasian treatment of these problems, he would have said, because that treatment was on lines with which he was very much out of sympathy. This lack of sympathy arose for a variety of reasons.

First, there was no obvious practical use in this way of setting things forth. "Professor Edgeworth's plan of representing U and V as general functions of x and y," he said in relation to a different but not dissimilar problem, "has great attractions to the mathematician; but it seems less adapted to express the every-day facts of economic

life than that of regarding, as Jevons did, the marginal utilities of apples as functions of x simply."[2]   Marshall, who wanted to use his theory of demand for statistical applications, would tend to be impatient with the complications of the Walrasian functions, and probably felt that, in the first approximation at least, for suitably chosen commodities, they could be neglected.

But, secondly, there was a deeper reason which he himself has set forth very explicitly.   On his view the important problems of theory were not problems of describing stationary equilibrium—which was altogether too easy — but rather problems of describing change and development.   And while, doubtless, here the statical method was appropriate enough, nay, even indispensable, mathematical methods of presentation tended to become intolerably complex and — what is more to the point — to emphasize the wrong factors.   "It would be possible," he said — obviously with Walras and others of that school in mind, but, with characteristic reticence, omitting derogatory explicitness — "to extend the scope of such systems of equation . . . . and to increase their detail, until they embraced within themselves the whole of the demand side of distribution.   But while a mathematical illustration of the mode of action of a definite set of causes may be complete in itself, and strictly accurate within its clearly defined limits, it is otherwise with any attempt to grasp the whole of a complex problem of real life, or even any considerable part of it, in a series of equations.   For many important influences, *especially those connected with the manifold influences of the element of time* (my italics), do not lend themselves easily to mathematical expression: they must either be omitted altogether, or clipped and pruned till they resemble the conventional birds and animals of decorative art.   And hence arises a tendency towards assigning wrong proportions to economic forces: those elements being most emphasized which lend themselves most easily to analytical methods."[3]   He goes on to say that nevertheless "in discussions written specially for mathematical readers it is no doubt right to be very bold in the search for wide generalizations."   But the general disinclination is writ large — and with Marshall it cannot be ascribed to any lack of aptitude for this kind of analysis.

I am not here concerned to defend this attitude.   I am concerned only to explain it.   I think myself that it was a great handicap for those of us who had to learn our economics from Marshall's *Principles* that he left so much unstated in the way of statical foundations.   We

2. *Principles*, 8th ed., p. 845.
3. *Op. cit.*, p. 850.

have had laboriously to rediscover for ourselves theorems and implications with which he was completely familiar all the time; and this has taken up too large a part of our lives. In this respect at least his passion always to be intelligible to businessmen and practically helpful to the world at large, has been a hindrance rather than a help. But I do think that, whether he was right or wrong in this respect, the underlying attitude which I have tried to make explicit, provides quite a sufficient explanation for the absence of much reference to the Walrasian system. He felt no personal debt in respect of those propositions of Walras with which he was in harmony: and for the rest I suspect that he definitely felt that the emphasis was either superfluous or in the wrong direction. The hypothesis of lack of generosity or of a desire — Schumpeter actually goes out of his way to suggest this — to "uphold the national tradition" against the wretched foreigners,[4] seems quite uncalled for.

What I think is perhaps even more surprising is Schumpeter's failure here to perceive more clearly the true nature of Marshall's contribution. In estimating the relative merits of Walras and Marshall — it is he who started this hare, not the reviewer — it is not enough to compare their achievements and priorities in the analysis of stationary systems. On this plane, perhaps, there is not much between them; as we have seen there is no doubt that Marshall had all the main theorems and some of the priorities. The point is rather that Marshall did so much more. The characteristic features of the Marshallian system, the analysis of the time element, the doctrine of short and long period normal price, quasi-rents, the dynamic theory of money and credit, the study of growth and decline, are all in a sphere which begins, so to speak, where the Walrasian system tends to leave off. Marshall himself protested against the notion that the central idea of the *Principles* was statical;[5] his claim was that it was "concerned throughout with the forces that cause movement"; and that "its keynote" was "that of dynamics, rather than statics"; and this claim was surely justified.

It might have been thought that this was a feature which would have made special appeal to Schumpeter, whose own central contribution had been so much in the theory of development: and there are passages, notably where he deals with Marshall's theory of monopoly, where such an attitude emerges. But on more fundamental conceptions, there was, of course, a very deep cleavage. It is not true, as Schumpeter seems at one place to suggest, that Marshall attached no

4. *Op. cit.*, p. 840.
5. *Principles*, 8th ed., p. xix.

significance to the innovating function of entrepreneurship:[6] quite the contrary indeed. Book VI and the historical sections, both of the *Principles* and of *Industry and Trade*, pay great attention to such factors. But viewing things in the large and not abstracting from the facts of fluctuation, he was inclined to emphasize the continuity of things — *Natura non facit saltum* was the motto of the *Principles* — whereas we know that Schumpeter's conception of the world involved much more of discontinuity and seismic convulsions.

VI

Part V, which is supposed to deal with the most recent developments, is a mere fragment — a series of first drafts which doubtless would have been completely remodelled had their author lived to finish them. Indeed, if the aesthetic unity of the book were the criterion, I should say that their inclusion was a mistake; though as a series of indications of the trends of Schumpeter's thought they were possibly worth preserving. The Introduction consists largely of the syllabus of a rather elementary course delivered at Mexico City. Chapter 2, on "Developments Stemming from the Marshall-Wicksell Apparatus" breaks off almost before it has begun. Chapter 3 on "Economics in the 'Totalitarian' Countries" seems almost pointless for the little it manages to scrape together; and Chapter 4, if it were intended to make very high claims for the progress of dynamics and business cycle research, falls considerably short of achieving its object. This is especially true of the short section on econometrics; after all that has been said earlier in the book about the prospect of placing economics on a more exact basis by research of this type, this jejune report has a distinct effect of anticlimax.

The only chapter which has any pretensions to move on the same level as the rest of the book is Chapter 5 on "Keynes and Modern Macroeconomics." This chapter is as interesting for its overtones as for its explicit content. It is clear that the spectacle of much contemporary Keynesianism inspired in Schumpeter a distaste that must have been almost physical. It is clear too that he was profoundly shocked by Keynes' occasional indifference to scholarly considerations, his eccentric, and frequently unjust judgments of his predecessors, and his obvious ignorance of much earlier work that was very relevant to his subject — deficiencies which, indeed, are inexcusable in smaller fry but which somehow or other do not seem to matter so much with him. His reaction to the man himself, too, was very ambi-

6. *Op. cit.*, p. 892–93.

valent; there seems to have been something in Keynes that alternately attracted and repelled him. But in this chapter he has made a great effort to be fair. The influence of Keynes is acknowledged, his stimulating effect on thought, especially the thought of the young, is emphasized. A handsome tribute is paid to his qualities as economic adviser. And while the assumptions of his theory are set out in a form in which an obvious light is thrown on its limitations, I do not think that any point that is made in this connection is unfair or invalid. Certainly the conclusion that the Keynesian paradoxes hold chiefly in a state of deep underemployment and that when the appropriate corrections are made for conditions of fuller employment, the theory becomes much more "Classical" in general content and implication, is one which would almost certainly have been accepted by Keynes himself — at least in his latest period.

None the less, I doubt if this is a chapter which will satisfy many. It certainly will not satisfy those who hold extreme views on either side: the extreme Keynesians will be content with nothing less than 100 per cent idolatry, the extreme anti-Keynesians with nothing less than excommunication with bell, book and candle. But the satisfaction of more middle views is very doubtful. The thing does not really come to life. There is little to which exception can be taken.[7] The general perspective is not seriously wrong. But it is as though the effort to be fair, the desperate resolution to say nothing that was not just and right, has used up all the author's vitality and penetration. The portrait is not incorrect. But the life has gone. This judgment may be too severe — after all, the chapter is an uncorrected fragment. Nevertheless, I cannot suppress the feeling that however extensive and finished the treatment, it would still have shown the same characteristics.

7. There is one very bad error of detail on p. 1171 where the outlook of the earlier pages of the *Economic Consequences of the Peace* is represented as anticipating, at least in regard to England, the author's views on saving — "the arteriosclerotic economy whose opportunities for rejuvenating venture decline while the old habits formed in times of plentiful opportunity persist." This is just not true. The outlook here is exactly the contrary — the war has shaken the institutions which fostered a high rate of saving, "The war has disclosed the possibility of consumption to all and the vanity of abstinence to many" (p. 22), the consequences for the future of capitalism may be disastrous. If anyone doubts this version, he may be referred to the following pronouncement from a *Manchester Guardian Reconstruction Supplement* (Aug. 17, 1922), written only shortly after. "In order to keep our heads above water, the national capital must grow as fast as the national labour supply which means new savings at the rate of £400,000,000 to £500,000,000 per annum."

VII

But I must not end on a note which might suggest any subtraction from the high praise of this book as a general performance which I gave at the beginning of this review.

The last time I met Schumpeter was on a river picnic in the middle thirties. He had turned up unexpectedly from the United States on the day of our annual seminar outing at the School; and he was immediately co-opted as an honorary member, so to speak, and pressed into joining the excursion. It was a lovely day in June; and, as we glided down the Thames between Twickenham and Datchet, I can still see him, cheerfully esconced in the prow of our ship, surrounded by the eager spirits of the day, Nicky Kaldor, Abba Lerner, Victor Edelberg, Ursula Hicks — Webb, as she then was, the master-organizer of the party — the four fingers and thumb of each hand pressed against those of the other, discoursing with urbanity and wit on theorems and personalities. So I conceive of this book, a splendid excursion down the river of time, with good talk and magnificent vistas.

LIONEL ROBBINS.

LONDON SCHOOL OF ECONOMICS

FULL EMPLOYMENT AND CONSTANT COEFFICIENTS OF PRODUCTION*

By MASAO FUKUOKA

I

The purpose of this paper is to present in an expository way an analysis of "structural" unemployment which cannot be adequately explained by prevailing Keynesian theory. It will be shown that the existence of this type of unemployment is rather closely related to the problem of the solubility of the Casselian equation system.

According to the Keynesian analysis, unemployment is due to a deficiency of effective demand, and hence full employment can be achieved through increasing the aggregate monetary expenditure. The truth of this conclusion seems to be beyond doubt if all factors are sufficiently available, and we might well understand how Lord Keynes was justified in presenting such a theory in England of the thirties, where the labor force and capital equipment were both crying for full utilization. However, the wind does not always blow in the same direction, and in the postwar economic situation the threat of inflation seems to have been a more urgent problem in most countries. Here again, if all factors are in bottleneck *pari passu*, the Keynesian remedy of cutting off expenditure may be *the* way to cure the disorder. Unfortunately the situation of war-shattered countries is, we must admit, far from the above. It has been frequently observed that in such countries capital equipment is seriously short relative to a rather abundant labor force, and a certain amount of unemployment results inevitably from the mere lack of equipment and materials to work with. If this is the case, a considerable volume of unemployment may possibly coexist with inflationary pressure, and here we have a special

type of unemployment which is radically different from that due to a deficiency of demand.[1]

Unemployment of this type is of great importance, for it occurs not only as a temporary postwar phenomenon, but also exists, more or less chronically, "in the backward, over-populated countries of the east, and, indeed, everywhere except among the most developed industrial nations."[2]

It is consequently very opportune that in recent years several economists have launched their efforts towards a formulation of such type of unemployment.[3] The present work represents another attempt to explore the theoretical implications of this subject.

1. The following illustration may help in understanding this situation. If all factors are available in a proportion such that full employment for each of them is reached at the same time, we may in fact regard them as a single "composite" factor and draw the well-known income-determination diagram as in Figure 1, where the dotted line shows the income level corresponding to full employment of the "composite" factor (see P. A. Samuelson, *Economics*, 2d ed., 1951, p. 287). In such a case, full employment of every factor can be achieved only by raising (or lowering) the level of effective demand $C + I$ (or $C' + I'$) to point F. In Figure 2, however, supplies of two factors (say, labor and capital) are not in
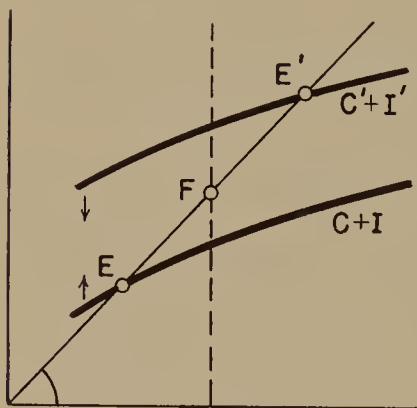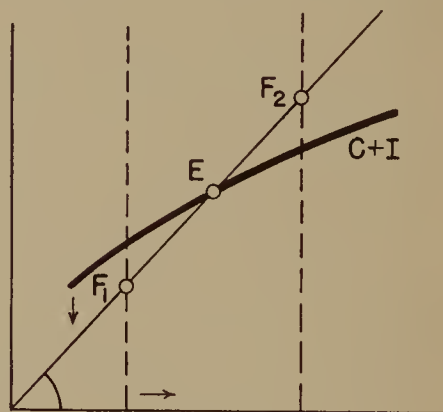
FIGURE 1   FIGURE 2

balance, and a bottleneck in one factor (say, capital) first causes a halt in production. In this case, despite the existence of an inflationary gap, the other factor (labor) must remain underutilized. The appropriate remedy then is not to raise $C + I$ to F_2 but to *lower* it to F_1 while at the same time to move the bottleneck line of the first factor itself as far as possible to the right in order to create room for raising effective demand.

2. Joan Robinson, *Collected Economic Papers*, p. 141.

3. For instance, see United Nations, *Measures of the Economic Development of Underdeveloped Countries*, 1951, and Ragnar Nurkse's recent book, *Problem of Capital Formation in Underdeveloped Countries* (New York: Oxford University Press, 1953), esp. Chap. II.

II

Before entering upon the analysis, a brief summary of earlier ideas may be in order.

A historical origin of the underlying thought could be traced as far back as the works of the classical economists. In his *Principles*, for instance, David Ricardo asserted that "with a population pressing against the means of subsistence, the only remedies are either a reduction of people or a more rapid accumulation of capital."[4] His disciple J. S. Mill concurred with him in this view and set forth the celebrated proposition that "industry is limited by capital," which implied that "there can be no more industry than is supplied with materials to work up and food to eat."[5] Karl Marx came a step closer to our present viewpoint by modifying the classical formulations on the following two points. First, he discarded the Malthusian mechanism of population adjustment and regarded the amount of labor rather as exogenously determined. Second, he separated the determinant of employment from the confusing "subsistence-fund" concept and attached it to the "machinery and materials" with which labor co-operates in the production process. Indeed, apart from the problem of effective demand, his famous "industrial reserve army" may well be interpreted as a description of the type of unemployment with which we are concerned here. For, from the purely technological point of view, it is nothing but the excess of available labor over the level of employment determined by the amount of capital and the fixed capital coefficient per unit of labor employed.[6]

The assumption of fixed technical coefficients was removed in the 1890's when the "modern" version of the marginal productivity theory was established by such exponents as J. B. Clark, Wicksteed, Barone, Walras and Wicksell. However, it may be noted that there were still some economists who were skeptical as regards the universal validity of the doctrine. For instance, Vilfredo Pareto, one of the most vigorous critics, observed that there might not be full substitutability between factors, and that some specific factors might be uniquely related to the product or to certain other factors. Thus, he

4. David Ricardo, *Principles of Political Economy and Taxation*, ed. Gonner, pp. 76–77.

5. J. S. Mill, *Principles of Political Economy*, ed. Ashley, pp. 63–64.

6. That Marx assumed the fixed coefficient of production is clearly observed in the following remarks: "To set in motion the part of social wealth which is to function as constant capital, or, to express it in a material form as means of production, a definite mass of living labor is required. *This mass is given technologically.*" (*Capital*, trans. Moore and Aveling, I, 669; italics are mine.)

said, " . . . there are other cases where one cannot compensate for the increase of one of the capital goods by the diminution of others. For example, in order to produce a given amount of silks, one requires an area of land to erect a factory, but, afterwards, even if one doubles this area, without increasing the other capital goods, the product will not be increased at all."[7] This is indeed the case of what Frisch called later "limitational factors."[8] Pareto's claim that the marginal productivity theory is "erroneous" is a rather deceptive proposition. Yet, from our present viewpoint, it is worthwhile noting that he showed how the production function might be restricted by certain technological conditions.[9]

The next important contribution appeared in the 1930's, when a group of Austrian, German and Scandinavian scholars, including Neisser, Stackelberg, Zeuthen, Schlesinger, Menger and Wald, raised questions on the solubility of the Casselian equation system, which assumed fixed coefficients of production and given availabilities of factors. Earlier critics such as Neisser and Stackelberg pointed out, though for different reasons, that the system does not always allow solutions satisfying the condition of demand-supply equality for all factors. Stackelberg in particular argued that on account of this

7. V. Pareto, *Cours d'économie politique* II (1897), section 717. Cf. also J. R. Hicks, "Marginal Productivity and the Principles of Variation," *Economica*, Feb. 1932, p. 86, n. 7. The present writer does not think this example so silly as does Professor Hicks.

8. R. Frisch, "Einige Punkte einer Preistheorie mit Boden und Arbeit als Produktionsfaktoren," *Zeitschrift für Nationalökonomie*, Bd. 3 (1931), p. 64.

9. Incidentally, it is interesting to note that Clark also regarded technical coefficients as fixed in the short run. However, he thought that, even if the amounts of factors were unchangeable, there would be, in the long run, a "qualitative" rearrangement of the given quantities of factors. In Clarkian terminology, a fixed amount of *capital goods* can employ only a relatively fixed amount of labor, while a fixed amount of *capital* can employ a widely variable quantity of labor. Thus, there are many capital goods that are already in use, "to which the bringing of one more workman would mean an excessive and uneconomical supply of labor, but there is no such limit to the number who can work with a fixed amount of capital, *if the forms of it can be varied to suit the number of the men.*" "If, therefore, capital is not limited in its forms, the labor that can use it is not limited in quantity," and "*an industrial society can in some way absorb any amount of labor.*" (J. B. Clark, *Distribution of Wealth*, pp. 114–15, the first italics are in the original, and the second are the writer's.)

To these remarks, let us add the following few words: It is true that capital equipment changes its forms in the long run according to the amounts of the other factors. But we should note that, if such a long time is considered, the amounts of the other factors would also change, and some of them (for instance, labor supply) might follow a somewhat autonomous behavior. Thus, even if sufficient time is given, there will be no such assurance that "an industrial society . . . can absorb any amount of labor." The dog will never catch the rabbit but always be behind it.

difficulty the system is forced to assume substitutability in production.   Though these observations may not be incorrect, it seems that the above mentioned incompatibility between fixed coefficients and full-employment solutions has its own merit because it clearly accounts for the possibility of some unused factors under the assumed circumstances.   Hence it is noteworthy that, following the suggestions of Zeuthen and Schlesinger, Abraham Wald extended the system in a significant way so that it includes both "scarce" factors and "free" factors (the former being those which are fully employed and the latter those which are available in excess of requirements) and proved the existence of a solution in the extended system.[1]   W. L. Valk of the Netherlands based his theory of "static" unemployment on this formulation of Wald and stressed that "when there are fixed coefficients, the scarcity of one factor may be the cause of oversupply of another," and "the most frequently observed and most interesting case is that in which there is too little capital to employ the entire laboring force."[2]   Related observations are also presented by K. J. Arrow and G. Debreu in their recent article.[3]

We may add as a fourth source the British economists' reflection on the nature of Keynesian "full employment."   In this connection, it must be admitted that Lord Keynes himself never left his flanks unguarded.   In Chapter 21 of his *General Theory*, we can find very interesting passages suggesting our present point of view.   He writes: "If there is a perfect balance in the respective quantities of specialized unemployed resources, the point of full employment will be reached for all of them simultaneously.   But, in general, the demand for some services and commodities will reach a level beyond which their supply is, for the time being, perfectly inelastic, whilst in other directions there is still a substantial surplus of resources without employment." "Since resources are not interchangeable, some commodities will reach a condition of inelastic supply whilst there are still unemployed resources available for the production of other commodities."[4]   Fol-

1. A. Wald, "Über einige Gleichungsystem der mathematischen Ökonomie," *Zeitschrift für Nationalökonomie*, VII (1936), pp. 637–70, trans. as "On Some Systems of Equations of Mathematical Economics," *Econometrica*, Oct. 1951, pp. 368–403.

2. W. L. Valk, *Production, Pricing and Unemployment in the Static State*, p. 58.

3. K. J. Arrow and G. Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, July 1954, p. 281.   Cf. also an excellent historical note, pp. 287–89.

4. J. M. Keynes, *General Theory*, pp. 296, 300.   In his *How to Pay for the War*, he went so far as to say: "The nature of unemployment today is totally different from what it was a year ago.   It is no longer caused by a deficiency of demand."   "It means that the Age of Scarcity has arrived *before* the whole of the

lowing this suggestion, Kaldor argued in his article of 1938 that, since "full employment" should be defined as the state where the output becomes inelastic and the point is reached when *either* equipment *or* labor is fully employed, there might be *either* "full employment of equipment" *or* "full employment of labor" and "only in particular instances . . . will it imply full employment of both." There he clearly pointed out that the productive capacity might be set by a scarcity of equipment despite the existence of unemployment of labor.[5] Joan Robinson made a similar remark when she stressed that "the question of *whether* full capacity [of equipment] *or* full employment [of labor] would be reached first" is "an important question" and in some cases "an analysis in which capacity is treated as the short-period bottleneck, restricting the possible rate of output at any moment, has more general application than one in which full employment sets the upper limit."[6] It is very suggestive that along this line of thinking she gave the following pertinent classification of unemployment which included the "Marxian" technological unemployment as well as the Keynesian depressionary unemployment: "Write A for the total of labour available, E for the actual level of employment and N for the amount of employment required to work the existing stock of capital at its normal capacity. . . . Now, A-E is the total of unemployment, A-N is the reserve army of labour, N-E is unemployment due to a slump."[7]

Finally there have been postwar observations by several economists. In his two interesting articles, Hicks analyzed the economic situation of war-damaged countries as follows. In those countries where the limit to productive capacity is set by severe shortage of equipment, marginal costs rise and hence real wages fall very rapidly as the technical limit is approached. Under such circumstances, an inflationary spiral of wages and prices may start far before the Keynesian "full-employment point" is reached, as a result of the pressure to maintain real wages. Thus, in a situation where such a "cost-of-living point" is reached before Keynes' "full-employment

available labor has been absorbed." " . . . Anyone who argues . . . that we are still in the Age of Plenty makes a mistake" (pp. 18–19).

5. Nicholas Kaldor, "Stability and Full Employment," *Economic Journal*, Dec. 1938, pp. 643–44.

6. Joan Robinson, *The Rate of Interest and Other Essays*, pp. 78–80.

7. *Ibid.*, pp. 110–11, n. 2. According to Professor Domar, Oscar Lange also tried to distinguish two kinds of unemployment. "The first, the Keynesian type, was common to advanced industrial countries and was caused by an excess of capital and a too high propensity to save. The second was usually found in undeveloped countries and was due to insufficiency of capital relative to population." Cf. this *Journal*, Nov. 1953, p. 562, n. 5.

point," there remains unemployment which cannot be solved by increasing effective demand.[8] A more or less similar phenomenon has been observed by Kalecki,[9] Simpson[1] and many others.[2] They all recognized that in the war-damaged countries unemployment was due to quite a different cause from that given by the prevailing Keynesian theory and, trying to find their way, opened the door to fresh consideration of the above ideas.

III

Proceeding from these sources, let us now begin our analysis by inquiring why the Keynesian remedy cannot lead to full employment of *all* factors. At first sight, one might be tempted to say that it is mainly due to the unevenness of the availabilities of factors. For, unless the given availabilities are such that every factor reaches full employment simultaneously, a bottleneck in one factor will necessarily appear in company with unemployment of other factors. However, this is apparently not sufficient to answer the query. If all factors are substitutable for one another, an increase in the price of the bottleneck factor will cause substitution of other factors for that factor. A further increase in effective demand will then tend to decrease unemployment of other factors and finally to result in overall full employment. Thus, for the Keynesian remedy to fail in removing unemployment, there must be a certain rigidity which prevents the above mechanism of free substitution. In order to make this point clearer, let us now take an extreme case in which all factors are strictly complementary, i.e., technical coefficients of production are absolutely fixed. In such a case, as soon as one particular factor hits a ceiling, it is no longer possible to increase output by increasing

8. J. R. Hicks, "World Recovery after the War," *Economic Journal*, June 1947, and " 'Full Employment' in a Period of Reconstruction," *Nationaløkonomisk Tidskrift*, 1948 (reprinted also in a Japanese journal, *Rironkeizaigaku*, Jan. 1951).

9. M. Kalecki, "Three Ways to Full Employment" in *The Economics of Full Employment*, pp. 43–44.

1. E. S. Simpson, "Inflation, Deflation and Employment in Italy," *Review of Economic Studies*, Vol. XVII (3), No. 44 (1949–50), pp. 203–25, esp. pp. 214 ff.

2. For instance, Vittorio Marrama, "Some Observations on Professor Lange's Analysis," *Economica*, May 1949, p. 126. Also F. Coppola d'Anna, "E'possibile una politica di 'full employment' in Italia?" *Previdenza Sociale*, July–Aug. 1946, which is cited in Marrama's review mentioned above, p. 125, n. 1.

As for Japanese contributions, I should like to mention Professor Hisao Kumagai's article "Jinko, Shihon oyobi Koyo," *Shogaku-ronshu*, Vol. 21, No. 3 (1952). The same thought was expressed a few years earlier in Professor Takuma Yasui's lecture at the Oriental Economist, which was entitled "Shotoku-kettei no Riron." The present paper was first inspired by this lecture.

expenditure. If demand is further increased, the only result will be a cumulative rise in prices, and *relatively abundant* factors will remain unemployed despite the inflationary pressure.

It should be noted that such a situation is different not only from the Keynesian type but also from that of the neoclassical theory which, assuming flexible substitution, asserts that unemployment can be cured through a change in (relative) prices of factors. Indeed, if unemployment of this type cannot be cured by increasing effective demand, neither can it be remedied by a neoclassical price reduction. Even if the price of any unemployed factor drops to zero, the factor will still be redundant without any increase in the amount of the scarcest factor.

The above consideration shows us what will be crucial in analyzing the outlined type of unemployment. The first thing to be done is obviously to specify in a certain way the nature of the production function. As a matter of convenience, we shall assume in the following that each industry has the early-Walras-Cassel-Leontief type of production function, i.e., that each uses factors in a fixed proportion although this may vary from industry to industry.[3] This assumption of fixed coefficients is not so far removed from reality as is sometimes thought. As Kaldor pointed out, "it is a significant fact of modern technique that *given the types of equipment in existence* there is a relation of strong complementarity between equipment, labor and raw materials."[4] Therefore, in so far as a change of technical coefficients is actually observed, it should be regarded as a dynamic transition from one set of coefficients to another rather than as a smooth substitution along the static schedule. Secondly, we shall assume that production is always carried to the limit of physical potentialities. This neglect of the Keynesian discipline should not be interpreted to mean that the present writer is unaware of the importance of "the theorem of effective demand." It is a deliberate device of abstraction to focus our attention on another cause of unemployment. As a matter of fact, if there is a deficiency of effective demand and the economy produces less than its potentialities, both Keynesian unemployment and our type will exist side by side, and unemployment will be found among all factors. Thus our assumption implies that the economy has already reached the situation where at least one particular factor is fully employed.

3. Though this assumption is very helpful in clarifying our analysis, it is not indispensable to the explanation of structural unemployment. For, as Neisser observed, even if the coefficients are allowed to change, there is still no reason why they should always match the relation among availabilities of factors.

4. Kaldor, *op. cit.*, p. 643.

Starting out with a simple model, let us consider that any one economy produces only two kinds of products. Denote by x_i and X_i the total and net outputs respectively of the ith product, and by a_{ij} the amount of the ith product needed for producing one unit of the jth product. Then, since the total of any good is allocated as final net output or as intermediate inputs, we can set up the following system of equations:

$$\begin{aligned} x_1 &= a_{11}x_1 + a_{12}x_2 + X_1 \\ x_2 &= a_{21}x_1 + a_{22}x_2 + X_2. \end{aligned}$$

(1)

This is a simple Leontief input-output system for two products, and can be solved in terms of the net products as

$$\begin{aligned} x_1 &= b_{11}X_1 + b_{12}X_2 \\ x_2 &= b_{21}X_1 + b_{22}X_2. \end{aligned}$$

(2)

where b_{ij} represents the x_i required to produce one unit of X_j. Now suppose that there are two basic factors 1 and 2, say, labor and capital, which cannot be produced within the above system. Let a'_{1j} and a'_{2j} be their amounts needed for one unit of the jth product. Then total requirements for these factors are respectively written as

$$\begin{aligned} a'_{11}x_1 + a'_{12}x_2 \\ a'_{21}x_1 + a'_{22}x_2. \end{aligned}$$

(3)

Substituting (2) into (3) and arranging terms, we can now express these requirements in terms of the net products

$$\begin{aligned} A_{11}X_1 + A_{12}X_2 \\ A_{21}X_1 + A_{22}X_2. \end{aligned}$$

(4)

where any A_{ij} represents the total ith basic factor needed to produce one extra net unit of the jth good. If we further denote by R_1 and R_2 initially given endowments of these factors, we are finally led to the familiar Casselian demand-supply equations for factors

$$\begin{aligned} A_{11}X_1 + A_{12}X_2 &= R_1 \\ A_{21}X_1 + A_{22}X_2 &= R_2. \end{aligned}$$

(5)

To be more general, we should add an inequality sign to each of the equations to get

$$\begin{aligned} A_{11}X_1 + A_{12}X_2 &\leqq R_1 \\ A_{21}X_1 + A_{22}X_2 &\leqq R_2. \end{aligned}$$

(6)

These last relations are the final all-important constraints to be imposed on the economy.

Now, under the assumptions made, our economy is supposed to produce as much output as physically attainable under the restriction of given resources.   However, in the present case of more than one product, what is the exact meaning of this "as much as"?   The prompt answer is the by-now well-known "efficient locus" (Koopmans) or the "production-possibility function" (Samuelson).   We may say that the economy maximizes the net output of the first product X_1, subject to a prescribed net output of the second product X_2 and to the linear inequality constraints (6).   It is obvious that the value of X_1 is determined by the following Leontief production function

$$(7) \qquad X_1 = \text{Min}\left(\frac{R_1 - A_{12}X_2}{A_{11}}, \frac{R_2 - A_{22}X_2}{A_{21}}\right),$$

where symbol Min is read that X_1 is the minimum of the expressions in the parentheses.   Note that, with R_i and A_{ij} being given, there will be a different attainable X_1 for each prescribed X_2.   The resulting production-possibility curve or "the best menu from which to choose" can be drawn as a heavy-black locus A_1CB_2 in Figure Ia or A_2B_2 in Figure Ib.
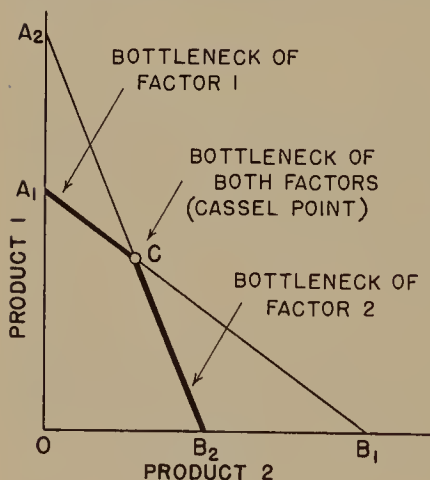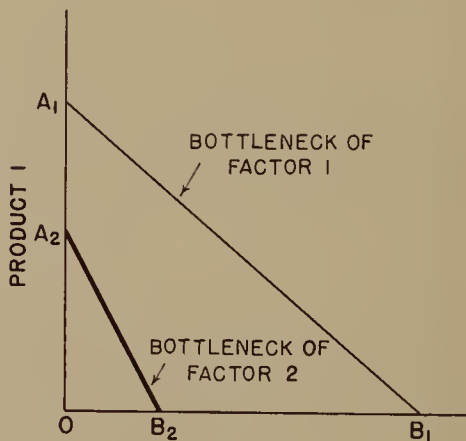
FIGURE Ia FIGURE Ib

It is now our task to examine what this analysis implies for the present problem of unemployment.   As can easily be seen, lines A_1B_1 and A_2B_2 in the above diagram represent the exact full-employment or bottleneck equations of each factor, i.e., they are $A_{11}X_1 + A_{12}X_2 = R_1$ and $A_{21}X_1 + A_{22}X_2 = R_2$ respectively.   The production-possibility curve is derived by following whichever line is inside, since the scarcest factor is always the bottleneck in production.   It is clear then that, if the production-possibility curve is to include a full employment of

both factors, the two bottleneck lines must intersect in the positive quadrant of the products. Only if this is the case, as shown in Figure I*a*, is full employment of both factors *technologically* possible. But if, on the contrary, those lines have no intersection in the positive quadrant, as in Figure I*a*, i.e., if either of the two (say, that of labor) lies uniformly outside the other (that of capital), strict full employment is impossible by any substitution between products.[5]

This analysis can easily be translated from the product space into the factor space. In Figure II, two dotted lines show the efficient combination of two factors in each industry. Then, in order that

FIGURE II
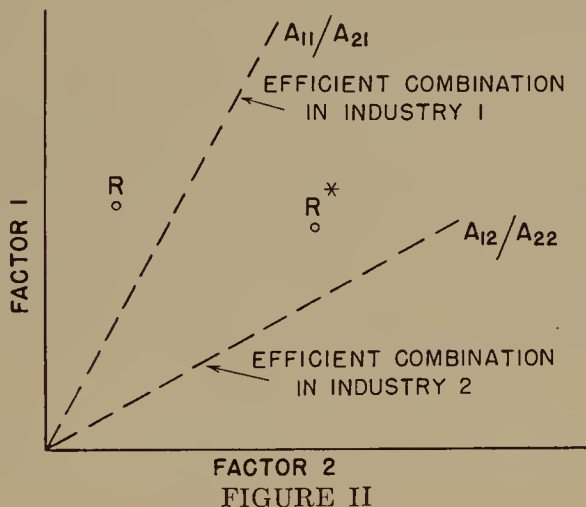
both factors be fully utilized, the point of given factor endowments, R, must lie within the area bounded by these dotted lines (like R^*), so that it can be attained by choosing proper positive outputs. If the point falls outside this area, then, as easily seen, full utilization of both factors is impossible without one of the outputs being negative, and hence a redundancy of one factor is inevitable. Thus, assuming $A_{11}/A_{21} > A_{12}/A_{22}$ without a loss of generality, we can represent the condition as

$$(8) \qquad \frac{A_{11}}{A_{21}} > \frac{R_1}{R_2} > \frac{A_{12}}{A_{22}},$$

5. At this point we must note that we are using the word "full employment" or "unemployment" in a rather peculiar fashion. From the Keynesian viewpoint, "full employment" is the state where the output cannot be increased any further and hence all the points on the efficient locus should represent "full employment" and "no unemployment." If we stick to this terminology, unemployment of our type should not be called "unemployment." Perhaps "redundancy" or "surplus" would be a better word.

which is easily checked to be the condition under which the solution of the strict Casselian equations (5) yields positive outputs. As a matter of convenience, such a solution will be referred to hereafter as the Cassel intersection point.[6]

IV

Adopting a purely technological point of view, we have shown above that over-all full employment could be attained only if the efficient locus contains the Cassel intersection point. However, as already stated, the efficient locus is nothing but "the best menu from which to choose"; hence, even if the "menu" contains the intersection, whether or not it is actually chosen will still depend on other factors than technology. In the present section, we shall introduce the concept of prices, and indicate that, even if the Cassel point exists, its realization via the market mechanism is still subject to a very restrictive condition.

The behavior of our economy has been described above as follows:

Maximize X_1, subject to any prescribed X_2, and subject to

(9)
$$A_{11}X_1 + A_{12}X_2 \leqq R_1$$
$$A_{21}X_1 + A_{22}X_2 \leqq R_2$$

and

$$X_1 \geqq 0, \; X_2 \geqq 0.$$

Now, introducing prices of two products P_1 and P_2, we may restate it in a slightly different fashion:

Maximize the (net) national product

$$Z = P_1X_1 + P_2X_2,$$

subject to

(10)
$$A_{11}X_1 + A_{12}X_2 \leqq R_1$$
$$A_{21}X_1 + A_{22}X_2 \leqq R_2$$

and

$$X_1 \geqq 0, \; X_2 \geqq 0.$$

6. From Figure II it is clear that the type of unemployment under consideration will not exist if there is at least one "product" which uses exclusively that factor. Indeed, if "leisure" is regarded as an alternative which competes with productive employment, there would always exist necessary complements for unemployable surplus. (See J. A. Schumpeter, *History of Economic Analysis*, New York: Oxford University Press, 1954, pp. 1010, 1012.) This seems to explain why unemployment of this type is liable to be "disguised" in so many cases.

If these prices are interpreted as hypothetical parameters, there is essentially no difference between these descriptions. Indeed, setting various price ratios parametrically and maximizing Z under each of them will leave us with the same production-possibility locus that we have derived before. Thus, in the case of Figure Ia, specifying each of the following price ratios:

(i) $P_1/P_2 > A_{11}/A_{12}$
(ii) $P_1/P_2 = A_{11}/A_{12}$
(iii) $A_{11}/A_{12} > P_1/P_2 > A_{21}/A_{22}$
(iv) $P_1/P_2 = A_{21}/A_{22}$
(v) $P_1/P_2 < A_{21}/A_{22}$,

we shall reach a maximum point of each of the following types

(i)   at the upper intercept, A_1
(ii) on the upper line segment, A_1C
(iii) at the Cassel intersection point, C
(iv) at the lower line segment, CB_2
(v)   at the lower intercept, B_2.
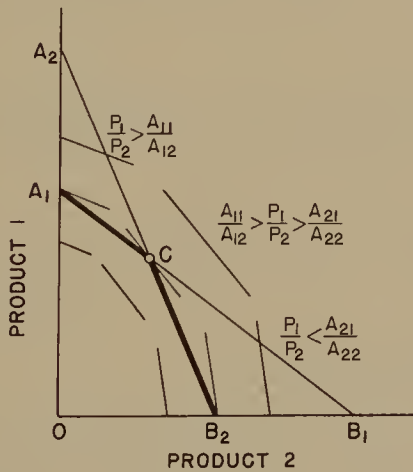
FIGURE III

However, it should be noted that which linear constraint is "binding" (i.e., which factor is in bottleneck) may now depend on which price ratio is specified. If the price ratio P_1/P_2 is given above A_{11}/A_{12}, the system is subject to only one binding constraint $A_{11}X_1 + A_{12}X_2 = R_1$. (See point A_1 in Figure III.) As P_1/P_2 falls slightly below A_{11}/A_{12}, it is switched into a new system with two binding constraints $A_{11}X_1 + A_{12}X_2 = R_1$, and $A_{21}X_1 + A_{22}X_2 = R_2$ (point C). A further fall in P_1/P_2 leaves the system unchanged until P_1/P_2 exceeds A_{21}/A_{22}. At this price ratio it becomes again subject to one binding

constraint, but this time to $A_{21}X_1 + A_{22}X_2 = R_2$ (point B_2).   This
suggests that the Cassel point will in effect be attained only if the
price ratio is given in a range A_{11}/A_{12} and A_{21}/A_{22}, i.e.,

$$(11) \qquad \frac{A_{11}}{A_{12}} > \frac{P_1}{P_2} > \frac{A_{21}}{A_{22}} \cdot,$$

Let us now explore the meaning of this condition further.   Since
(10) is a constrained maximum problem, we may construct the con-
ventional Lagrangian expression

$$(12) \qquad \phi(X, \lambda) = \sum_{j=1}^{2} P_j X_j + \sum_{i=1}^{2} \lambda_i (R_i - \sum_{j=1}^{2} A_{ij} X_j)$$

where λ_i's are Lagrangian multipliers, and maximize it with respect
to X_j's.   The first-order maximization condition gives us

$$(13) \qquad \begin{cases} \left(\dfrac{\partial \phi}{\partial X_j}\right)_{x_j = x_j{}^0} = 0 & \text{for positive (interior) } X_j{}^0 \\[3mm] \left(\dfrac{\partial \phi}{\partial X_j}\right)_{x_j = x_j{}^0} \leqq 0 & \text{for zero (corner) } X_j{}^0. \end{cases}$$

Since, if the Cassel point is to be the maximum, we must have the
exact equalities for both X_1 and X_2, we obtain from (12) and (13)

$$(14) \qquad \begin{aligned} \left(\frac{\partial \phi}{\partial X_1}\right)_{x_1 = x_1{}^0} &= P_1 - A_{11}\lambda_1 - A_{21}\lambda_2 = 0 \\[3mm] \left(\frac{\partial \phi}{\partial X_2}\right)_{x_2 = x_2{}^0} &= P_2 - A_{12}\lambda_1 - A_{22}\lambda_2 = 0. \end{aligned}$$

A remarkable fact to be noted is that condition (11) is now seen to be
the condition for a positive solution of the Lagrangian multipliers.
What is the economic meaning of these Lagrangian multipliers?
In order to answer this question, it is helpful to introduce another set
of Casselian equations

$$(15) \qquad \begin{aligned} A_{11}Y_1 + A_{21}Y_2 &= P_1 \\ A_{12}Y_1 + A_{22}Y_2 &= P_2 \end{aligned}$$

where new variables Y_1 and Y_2 are prices of two factors.   As is well
known, these equations indicate the equality between unit cost and
price for each product.   Now, just as we have extended (5) to (10),
let us extend (15) to the following form:

Minimize the national cost
$$Z' = R_1Y_1 + R_2Y_2,$$
subject to

(16)
$$A_{11}Y_1 + A_{21}Y_2 \geqq P_1$$
$$A_{12}Y_1 + A_{22}Y_2 \geqq P_2$$

and
$$Y_1 \geqq 0, \ Y_2 \geqq 0.$$

As $P_j - A_{1j}Y_1 - A_{2j}Y_2$ is the unit profitability of the ith product, the linear constraints imply that, if there is an excess profit, the market will force Y_i's to rise so high that profitability must not be positive anywhere. Further, if a product is not worth being produced, its profitability will have to be negative. The problem is thus to find the non-negative factor prices which minimize Z' under these constraints.

In terms of linear programming, (16) is known as the "dual" of the original problem (10), and its analytical treatment is quite analogous to that of the original one. We might well summarize it with diagrams similar to those in Section III. In Figures IVa and IVb, lines C_1D_1 and C_2D_2 indicate the competitive price-cost-equality

FIGURE IVa FIGURE IVb
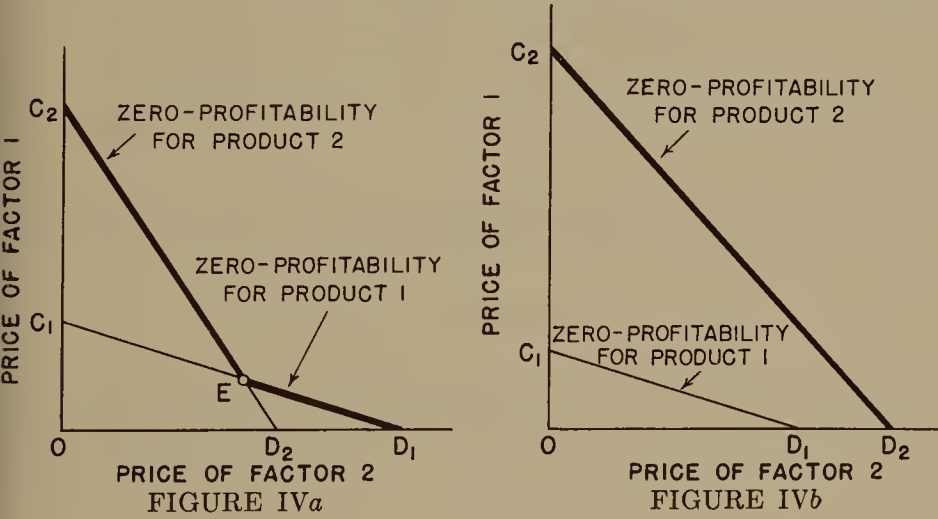
or zero-profitability equations for given P_i's, i.e., they are $A_{11}Y_1 + A_{21}Y_2 = P_1$ and $A_{12}Y_1 + A_{22}Y_2 = P_2$ respectively. Our constraints of nonpositive profits mean that we must always be above or to the northeast of every line. The boundary is shown as a heavy black locus C_2ED_1 in Figure IVa or C_2D_2 in Figure IVb. It is then obvious that, once any pattern of the R_i's is given, the corresponding best Y_i's are immediately determined.

The comparison of this with (13) and (14) will reveal that it is very natural to interpret the Lagrangian multipliers as equilibrium

prices of factors. Indeed, by a well-established theorem, we can verify that the solution of the minimization problem yields the Lagrangian multipliers for the constrained maximization problem. Thus the following conclusion presents itself: If the Cassel point is to be attained in equilibrium, system (15) must yield positive prices of factors, i.e., the two zero-profitability lines must intersect in the positive quadrant of the factor prices, as is shown in Figure IVa. Note the interesting dualism; the possibility of the Cassel point implies a positive solution of the first set of Casselian equations (5), while its attainability implies a positive solution of the second set of Casselian equations (15). It might be emphasized that in either case a failure of over-all full employment is caused by the fact that our variables are subject to fundamental non-negativity boundaries. This fact holds true whether they are outputs or prices of factors.

Incidentally, it will also be of interest to note that marginal productivity of any nonbinding factor is zero in equilibrium. For any ineffective constraint, the corresponding Lagrangian multiplier is naturally set equal to zero, which means in turn (according to what was mentioned above) that the equilibrium price of that factor is equal to zero. This will confirm our intuition that increasing any bottleneck factor will increase the maximum product, while increasing an unemployed factor will have no such effect until it begins to be binding.[7]

---

7. All this may be checked also in the following way: By a device of "slack" variables, we can rewrite our linear inequalities in terms of linear equalities

$$(1.1)\ \sum_{j=1}^{2} A_{ij}X_j + \epsilon_i = R_i\ (i = 1, 2) \text{ and } (1.1)'\ \sum_{i=1}^{2} A_{ij}Y_i + \pi_j = P_j\ (j = 1, 2)$$

where $\epsilon > 0$, $-\pi_j > 0$ $(i, j = 1, 2)$. Multiply each equation of (1.1) by Y_1 and Y_2 respectively and add them side by side. Similarly multiply each equation of (1.1) by X_1 and X_2 respectively and add them. Since the Duality Principle of Linear Programming tells us that

$$(1.2) \qquad \sum_{j=1}^{2} P_j X_j^0 = \sum_{i=1}^{2}\sum_{j=1}^{2} Y_i^0 A_{ij}X_j^0 = \sum_{i=1}^{2} R_i Y_i^0$$

holds in equilibrium for $X_j^0 \geqq 0$ and $Y_i^0 \geqq 0$, the following relations can easily be obtained

$$(1.3) \qquad \epsilon_i^0\ Y_i^0 = 0\ (i = 1, 2) \quad \text{and} \quad (1.3)'\ \ \pi_j^0 X_j^0 = 0\ (j = 1, 2).$$

The conclusions in the text follow immediately, i.e., (1) $X_j^0 > 0$ for both $j = 1$ and 2 implies $\pi_j^0 = 0$ for both $j = 1$ and 2. Hence, at the Cassel point, $A_{11}Y_1 + A_{21}Y_2 = P_1$ and $A_{12}Y_1 + A_{22}Y_2 = P_2$ must yield non-negative Y_1^0 and Y_2^0. "Non-genuine corners" being neglected, this must hold for positive Y_1^0 and Y_2^0. (2) $\epsilon_i > 0$ implies $Y_i^0 = 0$. Therefore, any redundant factor must have zero price in equilibrium.

To complete our discussion, it is now shown that from the description of our economy (10) a market-mechanism interpretation could be derived. We have noted above that the constrained maximum problem might be rewritten in terms of the Lagrangian expression (12). By a theorem of programming, we can further assert that, if we let (X^0, Y^0) be a saddle point of the Lagrangian expression, it follows that X^0 yields the solution to the maximization problem (10) and Y^0 yields a set of Lagrangian multipliers, which is none other than the solution of the minimization problem. This enables us to interpret our problem in the language of the theory of games, which may be described briefly as follows.

Our game of resource allocation is played by three persons: the "helmsman," the "manager," and the "custodian."[8] The helmsman determines prices of two products P_1 and P_2, and also sells to the custodian the available amounts of two factors R_1 and R_2. The custodian buys from the helmsman these amounts of the factors, sells them to the manager, and sets the prices of the factors. The manager buys the factors from the custodian, determines the outputs that can be produced under the limitation of the factors, and sells his outputs to the helmsman at the prices set by the latter.

Though there appear three persons in this game, the role of the helmsman is rather auxiliary, and we could regard the game essentially as a two-person (not necessarily zero-sum) game, the two players being the manager and the custodian respectively. The manager determines X_j's so as to maximize his gain (total profit).

$$(17) \qquad \pi^X(X, Y) = \sum_{j=1}^{2} \left(P_j - \sum_{j=1}^{2} A_{ij} Y_i \right) X_j \qquad \text{for } X \geqq 0,$$

while the custodian determines Y_i's so as to minimize his loss (total value of excess supply of factors)

$$(18) \quad - \pi^Y(X, Y) = \sum_{i=1}^{2} \left(R_i - \sum_{j=1}^{2} A_{ij} X_j \right) Y_i \quad \text{for } Y \geqq 0.$$

Note that the original product-maximization problem (and its dual, the cost-minimization problem, too) is now turned into a profit-maximization problem by the manager, and a market-clearing problem by the custodian.[9]

8. These names are borrowed from Professor Koopmans. Cf. "Analysis of Production as an Efficient Combination of Activities" in *Activity Analysis of Production and Allocation* (New York: John Wiley and Sons, 1951), pp. 93 ff. I also owe the following paragraphs to Professors Arrow and Hurwicz.

9. It is interesting to note that the clearing-of-the-market mechanism of demand and supply is expressed in terms of a minimization problem. Let excess supply of the ith good be Z_i. Then, for any given Z_i's, we can minimize the

Since expressions (17) and (18) are linear in X and Y respectively, we may further split each of them into their components. Thus, the manager of X can now be divided into the managers of X_j ($j = 1, 2$), each of whom maximizes his own profits ·

$$(19) \qquad \pi^{X_j}(X, Y) = (P_j - \sum_{i=1}^{2} A_{ij} Y_i) X_j \qquad \text{for } X_j \geqq 0 \ (j = 1, 2).$$

Similarly the custodian of Y can also be divided into the custodians of Y_i ($i = 1, 2$), who minimize

$$(20) \quad -\pi^{Y_i}(X, Y) = (R_i - \sum_{j=1}^{2} A_{ij} X_j) Y_i \qquad \text{for } Y_i \geqq 0 \ (i = 1, 2)$$

respectively.[1] One realizes now that the solution to our problem could be attained by an interplay of these decentralized interests. However, we should notice that, to whatever extent we might pursue the decentralization, we should end up with the same maximal configuration as the undecentralized process would attain. Thus, if the original configuration swerves from the Cassel point, the decentralized process would not tend to reach it either.

V

The preceding analysis will pave the way to our final discussion. Above we have assumed that the prices of the products P_1 and P_2 are set by the helmsman independently of outputs X_1 and X_2, and hence the expression maximized (maximand Z) is written in a linear form of X_j's. However, in the Casselian system, this is not so, and there are assumed certain functional relationships between prices and outputs. Let the demand functions be

expression $\sum_{i} P_i Z_i$ by increasing P_i if $Z_i < 0$ and by decreasing P_i if $Z_i > 0$
(unless P_i is not already 0), which is exactly the classical law of demand and supply. For this, see the decision-making of the "market participant" of Arrow and Debreu, op. cit., p. 275.

1. Dynamically we could set up the following sequence to describe the above process:

$$\begin{cases} \dfrac{dX_j}{dt} = 0 & \text{if } P_j - \sum_{i} A_{ij} Y_i < 0 \text{ and } X_j = 0, \\ \qquad = P_j - \sum_{i} A_{ij} Y_i & \text{otherwise;} \\[2ex] -\dfrac{dY_i}{dt} = 0 & \text{if } R_i - \sum_{j} A_{ij} X_j > 0 \text{ and } Y_i = 0, \\ \qquad = R_i - \sum_{j} A_{ij} X_j & \text{otherwise.} \end{cases}$$

The investigation of convergence of such sequence is another problem that cannot be discussed here.

$$(21) \qquad \begin{aligned} X_1 &= D^1 (P_1, P_2) \\ X_2 &= D^2 (P_1, P_2). \end{aligned}$$

On account of the homogeneity property, they can also be written in a single inverse expression

$$(22) \qquad \frac{P_1}{P_2} = N (X_1, X_2),$$

which means that the helmsman's relative marginal valuation is by no means constant but varies according to the variation of the outputs. Now we should take this relationship into consideration.

Though there might be several ways of approaching this problem, the clearest and most straightforward way would be to assume the existence of the social (or the helmsman's) utility function. Letting it be $U(X_1, X_2)$, (10) should be rewritten as:

  Maximize the social utility function $U(X_1, X_2)$,

  subject to

$$(23) \qquad \begin{aligned} A_{11}X_1 + A_{12}X_2 &\leqq R_1 \\ A_{21}X_2 + A_{22}X_2 &\leqq R_2 \end{aligned}$$

  and

$$X_1 \geqq 0, X_2 \geqq 0.$$

The corresponding Lagrangian expression becomes

$$(24) \qquad \phi (X, \lambda) = U (X_1, X_2) + \sum_{i=1}^{2} \lambda_i (R_i - \sum_{j=1}^{2} A_{ij} X_j).$$

Figures Va, Vb, Vc indicate the three sets of social indifference curves superimposed respectively on the production-possibility locus

FIGURE Va   FIGURE Vb   FIGURE Vc
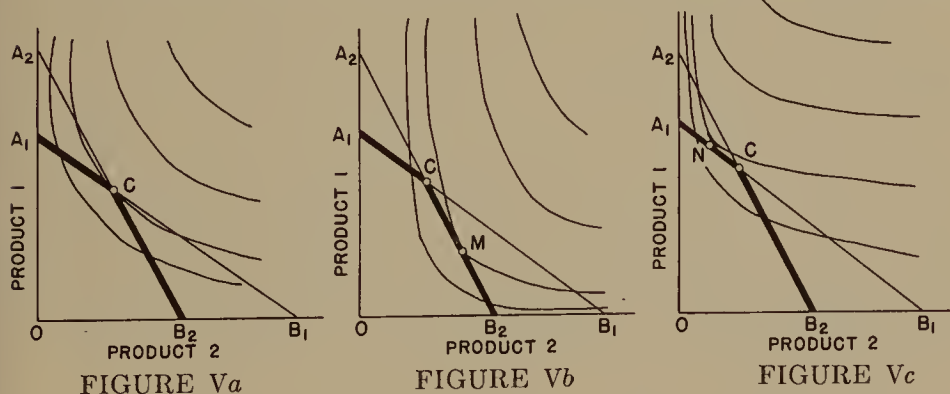
of Figure Ia. By looking at these diagrams, it is immediately evident that, if a final market equilibrium is to be reached at point C, the community indifference map must have a specific shape as drawn in Figure Va. If, instead, it is as drawn in Figure Vb or Vc, the final

equilibrium must necessarily contain a redundancy of one of the factors.[2]

If the expression $U(X_1, X_2)$ is substituted for $P_1X_1 + P_2X_2$, the description of the decentralized process must be altered slightly so that the decision-making of the managers is now based upon

$$(25) \quad \pi^{Xj}(X, Y) = \left(U_{Xj} - \sum_{i=1}^{2} A_{ij} Y_i\right) X_j \quad \text{for } X_j \geqq 0 \ (j = 1, 2)$$

where U_{Xj} denotes $\partial U/\partial X_j$. Now the helmsman announces U_{Xj}'s instead of P_j's, and re-announces the valuations according to the resulting outputs. In other aspects, however, the process is exactly the same as described before, and, if the maximal position is off the Cassel point, this process would again end up with some other point.[3]

VI

We may now summarize and conclude our analysis as follows: In the first place, we have seen that under the assumption of constant technical coefficients and given availabilities of factors, unemployment of some factors is inevitable unless these data satisfy a highly specific condition. Secondly, we have also seen that, even if technological data allow full employment of all factors, realization of that situation via the market mechanism is still subject to an extremely restrictive condition on the demand, or taste, side. Thus, it follows that, even if an economy is blessed with an effective demand large enough to induce maximum production, there will still be a possibility of unemployment, depending upon a specific structure of taste, technology, and availability of resources.

It is to be noted that this type of unemployment is fundamentally the result of an inconsistency in the given economic structure. Or, in more mathematical phraseology, it is owing to the fact that the solutions of the economic system do not yield positive outputs or positive

2. Formally this is the same problem as that of a consumer who is confronted by two budget constraints: a dollar and a ration-point limitation. Unless his indifference curves are of specific shape, either dollars or ration-points are redundant. Cf. Samuelson, *Market Mechanism and Maximization* (a privately circulated memorandum which will form part of a forthcoming RAND monograph on linear programming), Part II, pp. 13–14. Also see, *Foundations of Economic Analysis*, pp. 169–70, esp. n. 33.

3. The dynamical sequence should now be written as

$$\left\{ \begin{array}{ll} \dfrac{dX_j}{dt} = 0 & \text{if } U_{Xj} - \sum_i A_{ij} Y_i < 0 \text{ and } X_j = 0, \\[2mm] \quad\quad = U_{Xj} - \sum_i A_{ij} Y_i & \text{otherwise;} \\[2mm] -\dfrac{dY_i}{dt} = 0 & \text{if } R_i - \sum_j A_{ij}X_j > 0 \text{ and } Y_i = 0, \\[2mm] \quad\quad = R_i - \sum_j A_{ij}X_j & \text{otherwise.} \end{array} \right.$$

prices of factors. Therefore, if the phenomenon of unemployment is interpreted as arising from such causes, the only remedy would be to change the structure of the data in such a way as to bring the solutions into the meaningful positive orthant.

In the event that the physical situation allows for full employment but the pattern of demand prevents its attainment, a change in the latter might be considered as a required measure to be taken. Here, however, we encounter a difficult philosophical problem which stems from the philosophy of "consumer's sovereignty." Since, from this individualist viewpoint, people's preferences are to "count," it would be quite wrong for people to change their tastes so as to realize the technological possibility of full employment. Indeed, it may be that unemployment exists in the final equilibrium, but this is so just because people do not prefer the things which require all the available amounts of factors. Thus, as long as this philosophy is accepted, the final market equilibrium should always be desirable regardless of unemployment that it might contain. Now this will lead us to imply that, if by any chance the unemployed factor is important from the welfare or humanitarian point of view (labor, for instance) and the authorities of the economy should wish to decrease unemployment, they would necessarily be confronted with a discordance between their requirements. If the production-possibility locus is regarded as immovable, this discordance can be reconciled only through such a device as international trade. Otherwise a shift in the locus itself, through increasing the scarcest factor, will be the only way of surmounting the difficulty.

When full employment is technologically and physically impossible, the remedy should be sought in the direction of influencing (a) technology, or (b) the availability of factors. In the case of an overpopulated economy, technological improvements might consist in finding fruitful "capital-saving" techniques, since those that reduce the ratio of labor to capital will obviously be of no value unless they are accompanied by rapid economic development. However, it is frequently observed that in such an economy substantial technological improvements are of necessity "labor-saving." If this is so, a program for technological improvements will have to be twofold; to adopt labor-saving improvements of "industrial rationalization" in sectors where labor is adaptable, and, at the same time, to press forward the creation of new employment opportunities through as rapid capital accumulation as possible. Thus, to be fruitful, technological improvements will have to be accompanied by appropriate capital formation.

A policy of influencing the availability of factors should involve both the promotion of capital formation and the restriction of popula-

tion.   However, the latter remedy — including such measures as emigration, birth control and so forth — leads to somewhat extra-economic considerations, and does not seem to be so manageable (although not entirely impossible).   In view of the difficulties involved in population control, we may conclude that capital formation is the only desirable way of meeting the situation, especially in view of its importance in relation to technological improvements.

It should be noted that the above advocacy of promoting investment is based on a requirement quite different from that of Keynesian thinking.   In the latter, investment is needed because it generates effective demand which in turn stimulates employment by moving economic activity from the *inside* of the production-possibility curve towards its very frontier.   In our setup, however, the economy is supposed to be *on* the production-possibility frontier from the very beginning.   Investment is wanted just because it enlarges the capacity of capital equipment with which labor co-operates, in other words, it enables the production-possibility curve itself to shift uniformly outwards.   In the Keynesian case, such an enlargement of capacity may be regarded as a depressing factor relative to the future demand situation, whereas, in our case, the Keynesian remedy of expanding expenditure is bound to generate an inflationary pressure.   In the former, an increase in saving is considered positively harmful to the level of employment, while, in the latter, it is precisely what is needed for increasing employment without a burst of inflation.

As a matter of fact, most economies would involve both of these conflicting aspects to a greater or lesser extent, and which of the two would be dominant would depend on the situation confronting any particular economy.   It may be hoped that the theory outlined above will supply a helpful guide to one particular situation for which the Keynesian theory alone is inadequate.   Needless to say, however, in the final analysis the theory here developed should also be consolidated into a more comprehensive theoretical framework in which the influence of effective demand upon capital accumulation as well as that of investment upon effective demand are fully considered.

Apart from this point, our analysis is still subject to the following serious limitations: (1) capital equipment is regarded as if it were a "primary factor" the supply of which is given by nature, i.e., the dynamic analysis of capital accumulation is left out of account; (2) the technical coefficients are supposed to remain constant and hence the analysis of technological change is disregarded.   The development of these related points must await another occasion.

KEIO UNIVERSITY,   MASAO FUKUOKA.
SHIBA-MITA, MINATO-KU, TOKYO, JAPAN

SCALE OF OUTPUT AND INTERNAL ORGANIZATION OF THE FIRM*

By F. E. Balderston

I. Difficulties with current theory of output determination, 45. — II. Empirical tests of optimum scale: an example taken from the residential construction industry, 53. — III. Economies and diseconomies of scale, 67. — IV. Conclusions, 69.

I. Difficulties with Current Theory of Output Determination

How does the firm, under conditions approximating pure competition in its output market, find its best output level? This problem has long since been disposed of in economic theory, even to the matter of indicating stability conditions for the output equilibrium.[1] Yet the usability of the results as a realistic guide to the behavior of the firm remains somewhat in question.[2] Apart from the much-discussed problems of monopolistic competition and of strategy under oligopoly, there are two general sources of difficulty. The first concerns the precise nature of technical and other limitational factors that restrict the feasible output-level of the firm. The second has to do with the impact of uncertainty.

In Part I, we analyze important limitational resources — in particular, the executive organization of the firm and the amount of its money capital — and indicate shortcomings of current theoretical schema in portraying their effects. Uncertainty is considered as a characteristic problem in the determination of output and in the

1. Cf. P. A. Samuelson, *Foundations of Economic Analysis*, Chap. IV; and J. R. Hicks, *Value and Capital*, 2d ed., Chap. VI.

2. Cf. R. A. Gordon, "Short-Period Price Determination in Theory and Practice," *American Economic Review*, XXXVIII (1948), 287, for the contention that "the tool box of formal marginalism is not very useful in a business world characterized by as much *ignorance* and *uncertainty* (his emphasis) as do in fact prevail. These characteristics in turn stem primarily from two conditions: unending and unpredictable change and the existence of more 'directions of adjustment' (variables to be manipulated) than the business man can possibly handle in the manner assumed by formal theory." Gordon suggested that average-cost pricing, and a limited reliance on price changes for profit-maximizing adjustment, would constitute a closer approach to reality than the apparatus of marginal theory. The present paper was evoked by some of the same concerns that Gordon expressed, but an attempt is made to probe more deeply into the supply conditions for determining equilibrium output.

allocation of the limitational resources among production alternatives. In Parts II and III we present a linear programming model of the firm which appears to be more satisfactory than current theory in making the limitational resources explicit and in depicting their effects, and the effects of uncertainty, upon the output-level and internal organization of the firm. Empirical tests of the model are presented, with data drawn from the construction industry, to show what quantitative changes in profits, scales, and internal organization result from changes in the limiting factors.

A. *Limitational Resources and the Production Function.*

1. *Physical Capital and Executive Organization.* The "best" (i.e., profit-maximizing) output-level for the firm, under given prices on all input- and output-markets, is determined by the maximizing of a net revenue function (gross revenues minus costs) subject to technical limitations imposed by a production function. The most general treatments of the problem include little discussion of the nature of the constraining relation. Samuelson remarks that " . . . only 'inputs' [should] be included in the production function, and that this term [should] be confined to denote measurable quantitative economic goods or services."[3] At another point he says that the firm's total revenue must exceed total cost at least by the sale value of "owned" productive resources that are used up during the production period — otherwise the firm should go out of business.[4] Firms are normally considered to "own" two kinds of real productive resource: physical capital, and the talents and efforts of the entrepreneur himself, who might be thought of as coextensive with the firm. Only under remote contingencies is either of these resources likely to be placed on the open market. Should either of them, therefore, be considered to enter into a production function that is descriptive of output possibilities in the short run?

The ordinary requirements of a short-run output equilibrium are completely fulfilled even if these two kinds of limitational factor are omitted from the production function. Their short-run effect in producing both technical and cost limitations on output is indicated indirectly through increasing marginal requirements of some or all variable inputs, as output rises. So long as the theory is dedicated mainly to showing the effects on output of changes in input prices, and so long as the allocations of the constraining resources within the firm can be taken as the job of the entrepreneur, requiring no explana-

3. Samuelson, *op. cit.*, p. 84.
4. *Ibid.*, p. 88, note.

tion by the economic theorist, this device of limiting the production function to variable inputs only is quite satisfactory. The theory fails, both as an empirically-grounded explanation of the firm's behavior and as an explicit guide in executive decision-making, when any of the following conditions holds true:

(a) *The technical combinations of variable inputs are difficult or impossible to specify.* This can occur either because the entrepreneur is ignorant, or because the number of potentially usable combinations of inputs at each output level is so large that to specify them would be a practical impossibility within the statistical confidence limits desired for the production function, and within the budgetary limitations that exist both for the entrepreneur and for the scholar.

(b) *The adaptation of "owned" resources is not automatic.* Allocating fixed plant and equipment and fixed executive organization among alternative uses is in practice one of the most vexing top-level tasks within the firm. Applications of mathematical programming have indicated how physical plant and equipment can be dealt with explicitly.[5]

In general, however, economists have avoided explicit quantitative consideration of the executive resources of the firm as a limitation on output. Existence of a perfect market for supervisory personnel to whom some managerial functions could be delegated would not necessarily eliminate the problem. Even in long-run situations, complexity, and the burdens of management, might increase more than proportionately as scale increased.[6] Particularly for small firms, of course, the hiring of additional personnel at the executive level must be accompanied by an offer of ownership participation — i.e., by a dilution of the present entrepreneur's equity interest and a partial surrender of control.

Many writers have called attention to the fact that significant changes in the locus of control — whether in individual proprietorships, family partnerships, closely-held corporations, or large corporations controlled by a self-perpetuating managerial group — are generally considered disastrous; and it has often been pointed out that the

5. R. Dorfman, *Applications of Linear Programming to the Theory of the Firm* (Berkeley, University of California Press, 1951), and "Mathematical, or Linear, Programming," *American Economic Review*, XLIII (1953), 797–825.

6. Cf. E. H. Chamberlin, *The Theory of Monopolistic Competition*, 6th ed., Appendix B, pp. 247–49. Also, although the large firm may be better situated with respect to hiring possibilities in the executive market, the problem of *identifying* the "entrepreneurial interest" in the large enterprise impedes theoretical treatment of it. Methods of analyzing the nature of executive organization in such hierarchies are now receiving attention. Cf. C. I. Barnard, *The Functions of the Executive*, and H. A. Simon, *Administrative Behavior*.

desire to avoid loss of control exerts great influence on decisions with respect to the financial structure and the recruitment of executive personnel. These ideas were also borne out in the present author's study of a small group of residential construction firms. The small firm's top-level executive organization was absolutely limited in the short run, and alteration of the executive structure involved essential reorganization of the firm. Thus, there seem to be good empirical grounds for treating the executive staff as a fixed resource in the short-run theory of the firm.[7]

Economic theorists have also been baffled by the problem of specifying, with any precision, the tasks actually undertaken by executives. Even if it is known that the number of executives available to the firm is fixed, what limitations will a given distribution of numbers and specialized types of executive have upon the output-possibilities of the firm, and how will a new assignment of functions among these executives change the firm's output-level? To answer these questions, it is necessary to analyze the functional parts of the executive role in the firm and to discover how a contemplated change in the output-level may alter the burden of each type of executive work.

When the executive organization of the firm is viewed as a limiting factor, "make-or-buy" decisions assume a new form and more generally, all questions of vertical and horizontal integration become intertwined. The firm should distribute its executive staff so that both the output-level and the internal distribution of executive effort are consistent with the profit-maximum. In most instances a given amount and type of executive talent, combined with a given amount of money capital, result in a *unique* output-level and a *unique* distribution of executive effort which, together, will maximize profit.

2. *Working Capital Limitations.* Many statements of the theory of the firm are posited on the implicit assumption of unlimited access to circulating capital at a constant borrowing rate. Capital-rationing and an upward-rising supply-of-borrowed-funds curve are discussed by Hart[8] and by the Lutzes.[9] Makower and Baumol[1] also treat the

---

7. Limitations on the co-ordinating powers of the entrepreneur have been pointed out also as an inhibitor of indefinitely great size in the long-run equilibrium of the firm. Cf. N. Kaldor, "The Equilibrium of the Firm," *Economic Journal*, XLIV (1934), 60–76. This problem is also treated by R. H. Coase, "The Nature of the Firm," *Economica*, IV, New Series (1937), 386–405. Chamberlin, *op. cit.*, pp. 245–46, disputes Kaldor's assertion that long-run equilibrium depends necessarily upon the assumption that the entrepreneur is a "given" in the long run.

8. A. G. Hart, "Anticipations, Uncertainty, and Dynamic Planning," *Journal of Business of the University of Chicago*, XIII (1940), Part 2.

9. Friedrich A. and Vera Lutz, *The Theory of Investment of the Firm*.

1. Helen Makower and W. G. Baumol, "The Analogy Between Producer and Consumer Equilibrium Analysis," *Economica*, New Series, XVII (1950), 63–80.

limitation on money capital as a constraint on the firm analogous to the budgetary constraint in the theory of consumer behavior.

None of these discussions, however, shows how limits on executive resources interact with limits on money capital to produce a correct allocation of both at an optimum output-level. The Makower-Baumol formulation, which is seemingly the most complete of these discussions in this respect, was mainly devoted to a proof of the existence of "inferior factors." The entrepreneur maximized the value of a utility index, subject to a comprehensive transformation function which included both real and pecuniary magnitudes. Our objection here is not against the logic of that approach but against its applicability to a quantitative empirical problem. The entrepreneur's utility index is at the same level of abstraction as that of the consumer. The Makower-Baumol transformation function compounds the empirical mysteries of the production function with the difficulties of rationing working capital among alternative uses.

It hardly seems necessary to point out the relevance of money-capital limitations to a more realistic theory of short-run business behavior. Most commercial banks' lending practices are based upon rules which permit the firm to borrow up to a certain limit within a narrow band of borrowing rates. The borrowing limit is a function of the state of the money markets and of the size and reliability of the borrower. Since both of these may be considered given in the short run, a determinate limit on total working capital, in some definite ratio to the amount of the firm's own equity, will normally exist for each firm. The situation is, of course, complicated by the firm's opportunity to raise capital on a variety of bases: equity flotations; and secured and unsecured short-term and long-term loans having varying degrees of priority on the firm's assets in liquidation.

It may be possible to assume that a constant proportion of the unit cost of variable inputs can be financed through borrowings over a wide range of possible output-levels. Such is the case, for example, in residential construction firms when the conventional type of construction loan is utilized. The burden placed on the firm's own capital funds by each unit of output will then depend on this proportionality factor and on the extent to which the firm's own real productive resources are substituted for fully-fabricated purchased inputs.

The rate of turnover of the firm's own working capital is defined as the number of times, per (annual) production period, that the capital funds emerge in liquid form and can be used to finance new outlays for variable inputs. A highly variable element in the business

plan, the turnover rate will determine the *effective* size (in relation to the nominal or stated size) of the firm's working capital funds. The turnover rate will depend upon conditions in the money markets and the input and output markets, and upon the type of production schedule that is adopted by the firm. Any time-lag between completion and sale of a unit of output will be reflected in a lowering of the turnover rate and a consequent increase in the burden placed upon the firm's working capital position.

B. *The Subjective Climate of Decision-Making.*

1. *Uncertainty.* The accepted design of the theory of the firm suggests that the entrepreneur's main problem is to adapt to changes of input and output prices. We have indicated above, however, that resource limitations which bear upon the firm may in themselves pose a problem because it is necessary to ration these resources among alternative uses. It is partly an aesthetic question whether a model of the firm should deal with this question explicitly or implicitly; the present writer's preference is to place this problem of allocation in a central rather than in an implicit role. There are two operational reasons for doing so. First, it may be much easier to secure empirical estimates of the resource-limitations than of the myriad combinations of variable inputs. Executives in many firms will pose their output-determination problem in terms of the resource limitations. Also, the number of potential combinations is often less than in the case of the variable inputs. Second, when the resource limitations are made explicit, it is possible to treat much more fully the impact of uncertainty upon the internal operating patterns that are adopted by the firm.

Most suggested treatments of profit-maximizing under uncertainty involve correction of the relevant magnitudes to account for the degree of uncertainty and for the entrepreneur's attitude toward uncertain outcomes. Friedrich and Vera Lutz review the literature on this question and indicate how known probability distributions of uncertain outcomes can be reduced to certainty-equivalents.[2] There has been much debate on the relative merits of using the mean or the mode of the frequency distribution as the most probable value, on whether to correct this (since the expectations are not single-valued) by using the practical range or the standard deviation of the probability distribution, and on whether it is possible to combine proba-

---

2. Friedrich A. and Vera Lutz, *op. cit.*, Chap. XV. Among the contributors to this discussion are Hicks, Hart, and Lange. Cf. also, W. G. Baumol, *Economic Dynamics*, pp. 80–91.

bility-estimates of the original probability distribution (say, for the output price) with a series of probability-estimates for its (uncertain) parameters in order to derive a certainty-equivalent.

As Gordon points out,[3] much of the marginalism controversy concerns the realism of marginal comparisons under uncertainty; it may be the better part of valor to determine price and output on the basis of average cost. Fellner[4] shows how "safety margins" are maximized through the use of average-cost pricing; some compromise between this objective and the maximization of expected profits will be adopted, but the cumulative character of potential losses implies that safety margins will have the greater weight.

The fundamental difficulty with the certainty-equivalent approach is that each of the successive steps involved lends a progressively greater impression of unreality to the whole schema. It seems quite doubtful whether the method represents any approximation to the way in which people actually decide to do this or that. The underlying problem, of course, is to suggest through the model how uncertainty will reflect on the specific line of action taken by the entrepreneur, and how a change in his impression of the degree of uncertainty will alter his production plans and price-output equilibrium. In respect to this, Fellner's emphasis on safety margins has the merit of pointing out that there is a different rationale of price determination for protection of the firm's equity position than for the pursuit of maximum profits. If the resource limitations on the firm are made explicit, much more can be done to work out the effects of uncertainty.

It is suggested here that uncertainty has its chief effect upon the allocations of (limited) working capital and of (limited) executive time within the firm, and that these allocations will in fact control the output level that will maximize profit expectations subject to the given level of uncertainty.

Uncertainty may appear as a problem in the estimation of: (1) the parameters of a net revenue function; (2) the burden upon working capital incurred in producing each unit of output in each of various ways; (3) the rate of turnover of the firm's capital funds during the period; and (4) the burden upon the executive staff of producing each unit of output in each of various ways. It is suggested that the best course for the firm is to estimate most probable (i.e., mean or modal) values for the net revenue function, and then to account for the pres-

3. Gordon, *loc. cit.*
4. William Fellner, "Average-Cost Pricing and The Theory of Uncertainty," *Journal of Political Economy,* LVI (1948), 249–52.

ence of uncertainty by correcting (upward or downward, depending on the risk-preferences of the entrepreneur) the estimated values of the turnover rate and of the per-unit burdens on working capital and executive organization.

Besides determining the firm's best output-level under uncertainty, this method has the merit of showing how the firm's limited resources will be allocated for the specific purpose of avoiding the potential losses (or inviting the potential rewards) which, under uncertainty, have some chance of occurring. Thus the impact of uncertainty upon the firm's production program can be demonstrated in detail.

2. *Output-setting as a Choice Among Alternative Processes.* Explicit treatment of the executive resources of the firm may do considerably more than demonstrate some interesting properties of an important parameter of the firm's behavior. The modern theory of the firm concerns itself with the executive's purposeful actions, but it does not meet one of the prime characteristics of an acceptable theory of behavior: that the patterns of decision-making be approximated in the framework of the theory. The psychological setting for the making of executives' decisions ought logically to be defined by specialists in psychology and sociology and then be adapted to the uses of economic theory. We may draw upon the work of Katona[5] and upon case materials from our own investigations for some tentative observations:

(a) Much that is done in a business organization is essentially habitual and repetitive. These habitual modes of action may extend to things of considerable importance, such as a choice among suppliers of an important input. Only when habit becomes an excessively poor guide, or when strong external pressure goads the entrepreneur, does a close examination of alternatives result in a conscious decision which might be measured by standards of "rationality."

(b) The extensive use of habit is not good or bad — it is a reality. It can, however, be justified by the fact that close examination of alternatives requires very considerable expenditure of effort. The use of habit is an economy measure. The executive will upset an established way of doing things only when the potential returns — in terms of cost reduction, revenue increases, or possibly a more comfortable control position — can be expected to justify the disturbance caused and the effort required, or when a thorough examination of alternatives is forced on him and on the organization by strong internal or external pressures.

5. G. Katona, *Psychological Analysis of Economic Behavior* (New York, McGraw-Hill, 1951), Chaps. 9 and 10.

(c) So long as a habit-pattern remains in force, not much theory is required to explain its operation. Theory is needed to explain the ways in which a particular habit-pattern might be established in the first place, and to indicate the probable sources of disturbance.

(d) *A thorough examination of alternatives does occur only when they are in some sense psychologically distinct from one another.*

The analytical technique that is known as linear programming fits quite well the last, and most important, of the above requirements for a model of decision-making behavior in the firm, for programming involves the determination of a maximum (or minimum) position from amongst a finite set of distinct alternatives. Programming technique also makes possible the fitting of sparse data into an analytical schema — no mean advantage in the empirical examination of the problems of the firm.

II. Empirical Tests of Optimum Scale: An Example Taken from the Residential Construction Industry

A. *Some Cautionary Remarks About Linear Programming.*

Linear programming methods are used below to examine the process-alternatives that the firm may use in choosing its "best" output under circumstances of limited capital and limited executive organization. In the empirical cases to be presented, program solutions have been computed by means of the simplex method.[6]

Each process or activity in a program must be assumed to operate independently of every other, and each process must in principle be continuously divisible without loss of efficiency — operable at any non-negative level. When applied to the theory of the firm, linear programming thus requires the assumption of constant returns to scale.[7] As Chipman and Dorfman point out in their respective articles, it is not possible to avoid this assumption if the model that is adopted involves the maximizing of a linear relation subject to linear inequalities.[8]

B. *A Graphical Illustration of the Model.*

Our examples of scale-integration optima are taken from the residential construction industry. Residential construction firms are

6. A basic work in the field of linear programming is T. C. Koopmans (ed.), *Activity Analysis of Production and Allocation* (New York: John Wiley and Sons, 1951). Chaps. XXI and XXIII of that volume, both written by George B. Dantzig, give the proofs and certain computational techniques for the simplex method. Cf. also, Dorfman, *op. cit.*

7. J. Chipman, "Linear Programming," *Review of Economics and Statistics*, XXXV (1953), 101–17. Dorfman, "Mathematical, or Linear Programming," *loc. cit.*

8. See Section III, below, for further discussion of this problem.

generally small in terms of central organization. Many different patterns of operation coexist in the same market, both as to size of firm and type of internal organization. In the construction industry it is a frequent practice to subcontract a large proportion of the work and to reserve only a limited number of operations for performance by the firm's own labor force; but the amount and types of subcontracted work vary widely, and not necessarily in any regular relation to the scale of production. Finally, it is possible for the firm to rent all heavy capital equipment and to reserve its own funds for circulating use. This industry is thus a good one for observing the combined effects of limits on working capital funds and on executive organization without the complications of fixed plant.[9]

The linear programming approach to the determination of optimum scale and internal organization of the construction firm may be illustrated by a simple example involving the "one-man" firm.

The one-man firm is a business organization in which the entrepreneur supplies all equity capital and is the only executive capable of making decisions about price, output, and the disposition of the firm's resources. It is assumed, for simplicity, that the firm owns no physical plant and that the amount of the firm's own working capital is given in the short run.

We can distinguish between two types of executive effort: (a) supervisory or professional tasks to which the entrepreneur may assign himself or which he may hire out (these will be called *delegable* functions); and (b) basic decisions about organization, output, price and the disposition of resources (*nondelegable* functions).

It is assumed that the firm produces a single standard design of house and that (in this simple case) only two alternative methods of production (processes or activities) can be employed: the *integrated* method (x_1), which requires that the entrepreneur perform all delegable as well as nondelegable functions; and the *nonintegrated* method (x_2), which involves subcontracting all of the delegable functions. The integrated method is assumed to require a constant input (a_{11}) of executive time per unit of output, and the nonintegrated method, a (smaller) constant coefficient (a_{12}).

The firm is a pure competitor and is assumed to incur constant unit costs for labor, materials and rented equipment. These costs are financed by means of construction loans, so that the firm's own working capital is used for only two purposes: to purchase house-lots on

9. Cf. S. J. Maisel, *Housebuilding in Transition* (Berkeley, University of California Press, 1953), for a thoroughgoing description of the nature of the industry.

which to build, and to defray (in the case of the nonintegrated method only) the costs of hired supervision.

We may now write a net revenue function

(1) $R = b_1 x_1 + b_2 x_2$, which is to be a maximum subject to:

(2) $a_{11} x_1 + a_{12} x_2 \leqslant E$, the constraint on executive time, and

(3) $a_{21} x_1 + a_{22} x_2 \leqslant C$, the constraint on working capital.

The coefficients b_1 and b_2 are unit net revenues of the respective activities; $b_1 > b_2$, since outlays for hired supervision must be deducted from the gross margin only in the latter case. Also, $a_{11} > a_{12}$, for in the integrated method of production, the executive performs all delegable functions; and $a_{21} < a_{22}$, since no wages of supervision need to be paid in the former case. The program giving rise to maximum net revenue, R, is illustrated graphically in Figure I, for the situation in which both constraints are binding.

FIGURE I

Key:

    DF represents equation (2), the effort constraint.
    GH represents equation (3), the capital constraint.
      R is a net revenue line.
    OA is the number of houses produced by method "X_2"
    OB is the number of houses produced by method "X_1"
    GCF is the production-possibility bound.

C. *Organizational Choices: Activity Definitions and Coefficient Estimates.*

In testing this model of behavior, a fairly broad range of organizational choices is described. A set of thirteen activities (three of them "disposal activities," following the method of Dantzig) is therefore defined in Table I:

TABLE I

Activity:	Definition:
A_1	One-man firm engaged in site-fabrication of a single type of house for the mass market; executive performs all the delegable as well as the nondelegable functions. From the standpoint of use of the executive resource, this is the most "integrated" method of production.
A_2	One-man firm engages in site-fabrication of same type of house for same market as A_1, but executive delegates the function of on-site supervision by hiring a foreman or carpentry subcontracting organization, to be paid for supervision at a flat dollar rate per house.
A_3	One-man firm engages in site-fabrication as in A_1 and A_2, but in addition to hiring his on-site supervision, the executive buys house design and working drawings at a constant fee per house.
A_4	One-man firm engages in site-fabrication as in A_1, A_2, and A_3, but in addition to hiring on-site supervision and buying designs, the executive enters into an agreement with a real estate broker to sell his output at a fixed fee per house.
A_5	One-man firm buys prefabricated packages and produces houses to be sold in the mass market. Executive performs all delegable functions, as well as the nondelegable functions.
A_6	One-man firm buys prefabricated packages and produces houses to be sold in the mass market. Executive hires on-site supervision at a flat rate per house, but performs all other functions himself.
A_7	This is the activity of leaving idle some part of the executive's time.
A_8	This is the activity of leaving idle some part of the firm's working capital funds.
A_9	The firm has on its permanent payroll a supervisory assistant to the head of the firm. Houses are site-fabricated, are of a single physical type to be sold at mass-market prices. Executive and supervisor split up the performance of delegable functions, but executive reserves the making of all basic decisions for himself. No foreman is hired.
A_{10}	As in A_9, the firm has a permanent supervisory assistant to the executive, site-fabricates a single type of house in the mass

	market, and does not hire a foreman. The executive and the supervisor split the responsibility for the delegable functions in a different way than in A_9.
A_{11}	As in A_9 and A_{10}, the firm has a permanent supervisor and site-fabricates houses for the mass-market. A foreman is also hired, however, to superintend work on the site, and he is paid a flat rate per house.
A_{12}	This is the activity of leaving idle some part of the time of the permanent supervisory employee.
A_{13}	As in A_9, A_{10} and A_{11}, the firm builds houses for the mass market and has a permanent supervisory employee. However, it buys prefabricated packages. No temporary foreman is hired to supervise the work crews. Executive and supervisory employee split the delegable tasks between them.

It would be possible, of course, to broaden the range of choices by including owned capital equipment as a limiting resource, or by offering a selection of different market-sectors for output (i.e., price-classes of house), or by breaking down in greater detail the choices between doing the work with the firm's own labor force and subcontracting or purchasing fully-fabricated components. The present list of activities, however, gives some representation of the basically different methods of operation for small and medium-sized firms and will show some interesting empirical results.

For the disposal activities A_7, A_8, and A_{12}, the level of operation of the activity is measured as a percentage of the total supply of the respective limiting resources. Levels of operation for all other activities are measured in the number of houses produced.

Each activity requires a constant non-negative input of each resource per unit of output. Both the capital and the effort coefficients were estimated from case materials collected by the present author. The number of firms studied was small and not necessarily representative. The coefficients can be considered, at best, only rough approximations of "average-industry" coefficients, although the results of the model for different sizes of capital fund show remarkably close agreement with Maisel's data and conclusions concerning the structure of the house-building industry.[1]

1. Maisel, *op. cit.*, Chap. 1 and pp. 35–36, 71–72, 130–31. In the last of these cited passages, Maisel refers to limits on growth in the following words: "The average (large) firm has from one to four principals. . . . They make all of the primary decisions and many of the detailed ones. . . . In view of the personal nature of management in this industry, [expansion] can be done . . . . only by taking in additional principals. Adding principals means dilution of control and of returns. . . . Unless some personal force is driving toward increased size, expansion is unnecessary. If expansion is stopped, then the capital position improves and the risk of a total loss diminishes. . . . "

A consolidated set of effort coefficients is given in Table II.

TABLE II

UNIT NET REVENUES AND RESOURCE COEFFICIENTS FOR A
FIRM PRODUCING A SINGLE TYPE OF HOUSE FOR THE MASS MARKET

Activity: (1)	Unit Net Revenue: (2)	(3) Entrepreneur[1]	[Resource: (4) Permanent Supervisory Employee[1]	(5) Capital[2]
A_1	$1,400	9.85%	0%	20.00%
A_2	1,000	5.30	0	26.00
A_3	950	4.74	0	27.00
A_4	450	3.41	0	27.00
A_5	1,300	8.52	0	22.00
A_6	1,050	4.74	0	26.00
A_7	0	1.00	0	0
A_8	0	0	0	1.00
A_9	1,400	4.17	6.45	20.00
A_{10}	1,400	2.65	8.70	20.00
A_{11}	1,100	2.27	5.30	26.00
A_{12}	0	0	1.00	0
A_{13}	1,100	3.41	5.30	22.00

1. Coefficients stated in percentages of total available time.
2. Coefficients stated in percentages of a capital fund of $5,000, at turnover rate equal to unity.

Total available executive time was calculated as follows: of 365 days per year, subtract 1.5 days per week for leisure and 22 additional days as holidays, leaving 264 work-days. Estimates were made for the basic management functions and each of several delegable functions (i.e., design, purchasing, on-site supervision, and selling), in man-days per unit of output. These were then summed for each activity, and the sum was checked for correspondence with the facts available concerning the distribution of executive effort in case-group firms.

It would be possible to consider a partnership situation (if the partners were completely interchangeable and if both worked full time in the firm) by halving the percentage requirements for the entrepreneur in Table II. If a permanent supervisory employee is hired (as for activities A_9 to A_{13}), care must be taken to see that the executive and the supervisor do not duplicate the performance of functions.

As in the case of executive resources, it is possible to refine the functional breakdown of claims on the firm's working capital, per unit of output, to any degree that seems desirable. Possible inventory accumulations in advance of construction have been omitted from consideration. The firm has been assumed, however, to buy developed lots and to pay for subcontracted functions out of its capital funds,

and the latter items have been included in the capital coefficient for an activity if they occur in that activity. Even in the one-man firm, the amount of capital is potentially much more variable than the amount of executive time. Therefore, we have computed program solutions for sizes of fund ranging from \$2,500 to \$50,000, and the percentage coefficients of column (5), Table II, have been altered accordingly. The disposal activity for capital, A_8, has been assigned a coefficient of 1 per cent for all programs, however, so that the level of operation of this activity in the final program will indicate what percentage of available capital was left idle.

In the graphical illustration presented in part B of this section, nothing was said about the rate of turnover on working capital, and cash withdrawals by the entrepreneur during the production period were excluded by assumption. The first of these problems can be handled very simply. The actual size of the firm's capital fund, multiplied by the turnover rate, gives the effective size of capital fund available. The relevant set of coefficients may then be computed by multiplying the set of coefficients in column (5), Table II, by the ratio between \$5,000 and the effective size of fund.

The elapsed time of construction for project-built houses of the mass-market variety seems to vary between two-and-one-half and four months per house. Particularly when prefabricated packages are used, elapsed time is often reduced to the lower figure. If developed lots are purchased and if there is no delay in the sale of completed units, then it would seem reasonable to expect turnover rates ranging between 3 and 4.8 for a one-year planning period.[2]

Turnover rates are probably much more variable than this. During the period from 1947–51, except for part of 1949, builders in most metropolitan areas faced assured levels of demand for their houses. This fact had two consequences. First, builders were able to count on immediate occupancy of completed units by buyers. (In fact, many houses were sold in advance of construction, as buyers signed conditional sales agreements after looking at a single model house during the early stages of a project.) Secondly, short-term lenders were aware that hyper-boom conditions in the market for new construction reduced the risks of construction lending, and they pursued a policy of easy construction credit. The second of these considerations was strongly reinforced by the federal government's

---

2. Maisel (*op. cit.*, p. 40) says that " . . . the small (i.e., 1–9 houses per year) builder has a low turnover of his capital, his sales-equity ratio being approximately two-and-one-half to one, compared with a median ratio of five times per year for the large firm."

policies with respect to both the mortgage market and the money market generally. It would not be too much to say that a few builders managed, during this period, to engage in construction at annual rates ranging up to fifty or more houses without putting up more than token equity; i.e., turnover rates for such builders were probably so high that they were absolved from the effects of the capital constraint altogether.

Conditions in the money market and in the market for new construction can also lower the turnover rate very substantially. This fact is illustrated by the disappearance during 1951 of the extravagantly liberal short-term credits which some savings institutions had been willing to extend to builders in combination with long-term mortgage commitments, and by the emergence among commercial bankers of a tougher policy with respect to conventional construction financing. During periods of sharp recession, on the other hand, builders may find their annual turnover rates slipping to two, or possibly as low as unity. This downward movement would be caused not only by the tightening of short-term construction credit but by some increases in the time-lag between completion and sale of new houses. Application of the scale-integration model to a particular situation should always be based upon an explicit estimate of the turnover rate, corrected by the entrepreneur's attitude toward the degree of uncertainty in the market.

Particularly in the case where use of prefabricated house-packages is under consideration, it may be desirable to set up some models in which, for given resource limits of the firm, coefficients corresponding to a turnover rate of say, four, are adopted for the prefabricating activities A_5 and A_6, whereas turnover rates of two or three hold true for the site-fabricating activities.

The problem of cash withdrawals or payment of salary to a hired employee can also be dealt with in an approximate way. Assuming that a payment is made at the end of each month, let C be the firm's actual capital funds, S, the employee's annual salary (or the entrepreneur's annual cash withdrawal), r, the turnover rate, and M, the effective size of the capital fund. Then $M = r\,(C - (S - 78/144 \cdot S))$, where the fraction $78/144$ indicates the weighted-average number of salary-months during which the proposed salary payments will actually be available for use as working capital. Except where otherwise stated, no corrections for cash withdrawals are made in the program solutions that follow.

The net revenue function is linear (under the assumption of a purely competitive market, or the (weaker) assumption that, what-

ever the degree of imperfection, the entrepreneur is too uncertain to make any allowance for it by postulating a downward sloping demand curve and will merely add safety-factors to the activity-coefficients), and for each of the models encompassing activities A_1 to A_8 it will be written

$$R = 1{,}400\, x_1 + 1{,}000\, x_2 + 950\, x_3 + 450\, x_4 + 1{,}300\, x_5 + 1{,}050\, x_6$$
$$+ 0\, x_7 + 0\, x_8,$$

where R is in dollars of net revenue, and the x_i are the (unknown) activity-levels. If the firm has on its payroll a permanent supervisory employee, however, the revenue function is altered by a lump-sum subtraction of his annual salary, since that amount is independent of the amount of revenue produced by any activity. So far as the entrepreneur is concerned, of course, his objective is to maximize the *sum* of his wages of management and returns to capital, so that the above net revenue function should hold even if cash withdrawals are made.

Table II contains, as stated, a set of approximate average-industry coefficients. The relation between these and the application of the analysis to a specific firm in a specific situation should be briefly stated. The entrepreneur in such a firm may be expected to have certain strengths and certain weaknesses as compared with the industry average. He should therefore alter the effort-coefficients, before correcting for uncertainty, so as to reflect his own talents. The coefficients of the net revenue function that he uses should be the most probable unit values derived from an estimate of the local market situation of the firm. Finally, uncertainty should produce a stimulus to the entrepreneur to allocate extra working capital and extra executive effort to each unit of the contemplated set of activities. The algebraic size of these safety factors will depend on the degree of uncertainty felt by the entrepreneur and on the extent of his desire to avoid a possible loss-outcome.

D. *Empirical Results: Breaking Points between Types of Firm.*

1. *The Undercapitalized One-Man Firm.* The list of potential activities which the firm may operate is A_1 through A_8. With working capital funds of $5,000, no cash withdrawals, and a turnover rate of unity, this program has the optimal solution ($x_1 = 5$ houses; $x_7 = 50.5$ per cent of executive time left idle; $R = \$7{,}000$ of net revenue).[3] This optimal program holds for the situations in the table on the following page.

3. The size of expected net revenue is close to Maisel's estimates of $7,800 combined wages and profit for the median small builder, who produced between 1 and 9 houses per year. (Cf. Maisel, *op. cit.*, p. 36 and Appendix E.) Such

Capital Fund	Turnover Rate	Cash Withdrawals
$10,000	0.5	0
5,000	1	0
2,500	2	0
1,667	3	0
1,250	4	0
7,500	1	*
5,000	2	*
4,167	3	*
3,750	4	*

*Withdrawals totalling $5,000 annually are assumed to be made in twelve equal installments, paid at the end of each month.

The above list of situations in which the firm would have an *ex ante* net revenue of $7,000, leaving half the entrepreneur's time idle, is instructive, first of all, in showing the important impact on the small firm of the turnover rate on its working capital. During a hyper-boom period, the entrepreneur who could devote all of his capital funds to house-building could support a program of five houses with only $1,250 in working capital, at a turn-over rate of 4. During depressed periods, when the turnover rate might be as low as 1, an equally hard-hearted entrepreneur (from his wife's point of view) would require $5,000. If he withdrew funds from the firm on a monthly basis when the turnover rate was equal to unity, he would need $7,500 of working capital. The strong impact of cash withdrawals by the entrepreneur of the small firm is thus clearly emphasized.

It is the writer's view that the considerable quantitative difference in working capital funds required to sustain a given building program under different turnover rates helps to explain the high rate of entry into the industry during boom periods, and the high exit rate of small marginal firms during recessions. It should be noted that this difference shows up strongly even though no correction has been made in the revenue function for widening of unit profit margins during booms, and narrowing of the margin during depressed periods.

The internal organization of the firm under all of the situations for which this program holds true is quite simple. The entrepreneur turns his hand to every professional task: making the basic decisions (the nondelegable functions); doing the designs; purchasing; selling; and acting as his own foreman. Even so, he cannot apply all of his time to these professional functions because, under the given circumstances, the capital base of the firm is insufficient to permit his gain-builders were estimated to have about $7,000 of cash and cash-equivalent on hand at any given time, with current assets (including work in process) of $15,000. (*Ibid.*, p. 38.)

ing sufficient output volume to be fully employed at the executive level. It is likely to be true, of course, that the entrepreneur will dip down even further into the pool of work available, and thereby displace an hourly-paid carpenter or other employee. This will have the effect of increasing total net revenue by a little more than the amount of wages saved. That part of the firm's construction borrowings which is not paid out to the displaced laborer can be used to support other capital needs of the firm. If the entrepreneur can displace, with half of his time, half the time of a skilled carpenter who is paid at the annual rate of $4,000, then the capital funds of the firm will be effectively augmented by ($4,000 $\times \frac{1}{2} \times$ 78/144).[4] Therefore, if the head of the firm does employ himself fully by displacing an hourly-rate employee, the above results are altered to show substantially reduced working capital requirements in each of the situations noted above. For example, at a turnover rate of unity, with no cash withdrawals, the working capital needed to produce five houses is reduced from $5,000 to $3,920. In order to be plausible, the results of this model require that the firm work out a synchronized schedule of production and selling *during* the planning period, such that extraordinary peaks of working capital needs are eliminated.

2. *The One-Man Firm Having Just Enough Capital to Keep the Entrepreneur Fully Employed as an Executive.* The list of potential activities is A_1 through A_8. With working capital funds of $10,000, no withdrawals by the entrepreneur, and a turnover rate of unity, this model has the optimal solution (x_1 = 10 houses; x_7 = 1 per cent of executive time left idle; R = $14,000 of total net revenue).[5] As before, the actual size of the capital fund that is required varies in inverse proportion to the size of the turnover rate, so long as zero cash withdrawals are made. Where cash withdrawals occur on the same basis as in our first case, the required size of fund is $12,500 at a turnover rate of unity, $7,500 at turnover of two, and $5,833 at turnover of three.

The executive is 99 per cent employed in performing the nondelegable and all of the delegable managerial functions. There is no need, therefore, to make any allowance for the possibility that the

---

4. Maisel notes that 76 per cent of the owners of small (1–9 houses per year) firms were "working foremen who spent six to seven hours each day on the job as a carpenter and foreman." (*Op. cit.*, p. 35.)

5. Maisel reports that "as firms expanded into the 10–24 classification, their methods of operation began to change." (*Op. cit.*, p. 33.) And, "As volume increases, overhead structure also develops. . . . In the 10–24 completion group nearly a third of the owners operate as working foremen; somewhat more act as job superintendents, and full-time executives. . . ." (*Ibid.*, p. 37.)

executive would attempt to displace an hourly-rate worker with part
of his time. At the turnover rates specified, the amounts of capital
funds are still too small to permit the executive to concentrate his
whole time on the making of basic decisions and the other tasks that
are thought of as the nondelegable functions of the head of the firm.

The program that is specified in this case is a most important one
because it shows what capital funds are needed, under various turn-
over rates, to keep fully employed an entrepreneur who is able to do
his own work as foreman of the job. Most firms in this size-group
would probably be observed to require the larger capital funds (at
each of the possible turnover rates) that result from a policy of regular
and sizeable cash withdrawals; it is generally true that full-time heads
of firms in this industry depend upon the firm for ordinary living
expenses during the year.

3. *The One-Man Firm: Comparison of Site-Fabrication and Pre-
fabrication at Different Turnover Rates.* To the small builder, two
potential advantages in the use of prefabricated packages, as against
the conventional site-fabricating techniques, are that some managerial
functions are simplified, and that the shorter elapsed time required to
put up prefabricated units increases the rate of turnover on working
capital.[6] In order to demonstrate how this might work out in an
approximate way, let us consider a one-man firm having $5,000 of
capital funds and a policy of not making cash withdrawals. It was
shown earlier that such a firm would specialize on activities A_1 and
A_7. If the capital turnover rate is 2, the optimal program is ($x_1 = 10$,
$x_7 = 1$ per cent), producing net revenue of $14,000. Suppose, how-
ever, that the shorter elapsed time of building prefabricated houses
enables the firm to turn over its working capital *three* times per year
if it chooses the activities (A_5 or A_6) involving prefab packages. We
now solve a new program to test the possibility that the higher turn-
over rates for the latter activities will cause either or both of them to
replace A_1. The equation system is:

$$R = 1{,}400\ x_1 + 1{,}300\ x_5 + 1{,}050\ x_6 + 0\ x_7 + 0\ x_8,$$

to be maximized, subject to:

$$9.9\ x_1 + 8.5\ x_5 + 4.7\ x_6 + x_7 = 100\ \text{(the effort constraint)},$$
and $10\ x_1 + 7.4\ x_5 + 8.7\ x_6 + x_8 = 100$ (the capital constraint).
The solution to this program is ($x_5 = 10$, $x_6 = 3.04$), producing a
total net revenue of $16,190. Instead of performing all executive
functions himself, including on-site supervision, the entrepreneur even

6. Maisel (*op. cit.*, p. 292) notes that small builders might use prefabricated
packages to gain some advantages of both the small and the large firm.

hires a foreman to manage the work crew for three houses.   Total output has risen by 30 per cent and total revenue by 16 per cent as a result of the change in the program.

Another advantage to the small firm of using the prefabricated package is also evident here: opportunities to build in mass-market price classes, as against building in higher price ranges, are increased because the prefabricated package helps the entrepreneur to economize on his time and capital, per unit of output.

4. *The Effect of Further Increases of Capital Funds in the One-Man Firm.*   For the firm having $15,000 in capital funds with a turnover rate of unity in all uses, the optimal program is (x_1 = 7.4, x_6 = 5.6), which produces a net revenue of $16,500.   As before, for the cases in which no cash withdrawals are made during the year, the same program can be sustained with proportionately smaller amounts of capital for higher turnover rates.   Once again, a correction for cash withdrawals, however, would increase the size of capital fund required to sustain this program at each given turnover rate.

The program solutions for still greater sizes of capital funds, again with turnover rate equal to unity and a policy of no withdrawals, are as follows:

$20,000 fund: x_5 = 5.8, x_6 = 10.5,   revenue = 18,565
$30,000 fund: x_6 = 21.3, x_8 = 8.4,   revenue = 22,365.

There might appear to be a conflict between the first of these solutions (for the $15,000 capital fund) and the solution to the special problem of subsection 3, above.   In the latter, the optimal program consisted of activities A_5 and A_6, whereas in our present group of solutions, activity A_1 is not driven out until the capital fund reaches a level of $20,000.   In this group of solutions, however, the capital coefficients for all activities are gradually falling, *in the same proportion*, whereas the assumed conditions of the special problem produced a different capital coefficient for x_1 than appears in the first problem of this group.   This explains the more radical shift out of A_1 in the special problem than in this group of solutions.

It is still observable, however, that as the capital funds of the firm increase (or as turnover rates increase equally in all activities for the same size of capital funds), a shift toward use of the prefabricated package occurs.   This is occasioned by the fact that demands on executive time are lowest, relative to potential profits, in the activities associated with prefabricated units.   Furthermore, the use of prefab packages permits the executive to concentrate, in activity A_6, on only three key functions: the making of basic decisions; the

purchase of inputs (which requires less time here than in A_1 because a big part of the materials bill is secured in the package); and the selling of houses.

When the capital fund reaches $30,000 (or proportionately less for turnover rates higher than unity), the firm cannot absorb all of the fund in its optimal program. Only the effort constraint is binding. Capital funds of this magnitude give rise to the possibility of supporting a larger executive organization than exists in the one-man firm. (If the firm has been operating at a higher turnover rate than it expects to attain in the future, however, it may wish to hold liquid capital in reserve against a time of greater need, rather than expanding the central organization.)

5. *The Firm Having a Permanent Salaried Employee in Addition to the Active Entrepreneur.* In the program solutions which were demonstrated above, the executive of the firm had the option of securing, or doing without, a number of hired services. The following managerial functions were involved: design; on-site supervision of the work crew; and selling. In any case where the entrepreneur did hire these functions done, it was assumed that he paid a flat rate, per house, for the work on each type. Thus, the payments made could be reflected in the increased demands on the capital funds of the firm (except in the case of a real estate agent's fee, which was assumed to be collected upon receipt of cash for sale of the house, and therefore to require no drain on working capital.) Payments for these types of work also reduced the unit net revenues of the activities in which these appeared as delegated responsibilities.

When the firm takes on a permanent staff member at the supervisory level, however, his salary payments are a function of time, not of work on particular units of output. The method of accounting for these payments by altering the effective size of the working capital fund has already been described.

The set of activities involving the use of a permanent staff member is A_9 through A_{13}. Our first task, however, is to show whether it is worth the firm's trouble to employ a permanent man. Consider the following situation:

> Capital fund = $15,000; turnover rate is unity; neither the entrepreneur nor the supervisor is assumed to receive any cash payment during the production period; the firm may choose among activities A_1 through A_{13} on the basis of maximum profit; supervisor's annual salary is assumed to be $4,000.

The revenue function is: $R = -4,000 + 1,400\,x_1 + 1,000\,x_2 + 950\,x_3 + 450\,x_4 + 1,300\,x_5 + 1,050\,x_6 + 0\,x_7 + 0\,x_8 + 1,400\,x_9 + 1,400\,x_{10} + 1,100\,x_{11} + 0\,x_{12} + 1,100\,x_{13}$, which is to be maximized.

The solution to this program is: $x_7 = 37.2$ per cent of executive time left idle; $x_9 = 14.9$ houses produced; and $x_{12} = 2.9$ per cent of the supervisor's time left idle. Total net revenue is $16,900, after deduction of the $4,000 salary paid to the permanent supervisor.[7] Another program solution gives the same total net revenue, but keeps the entrepreneur at work a slightly larger proportion of the time. This solution is: $x_1 = 3.4$ houses; $x_{10} = 11.5$ houses; and $x_7 = 35.9$ per cent of executive time left idle. We presume that, for the same total net revenue, the executive would wish to permit himself as much leisure as possible, and that the first of these solutions is therefore preferred. Again, a larger total amount of capital will be needed to sustain the optimal program if cash payments to the entrepreneur and the supervisor are made during the year. Other programs — involving larger effective sizes of capital fund or larger numbers of supervisory employees — can be explored by methods similar to those used here.

III. Economies and Diseconomies of Scale

Many industrial processes are said to be subject to economies of scale, in that large aggregations of machinery must be brought together to satisfy minimum requirements of technical and economic efficiency.[8] Even within the industry to which our present investigation refers — residential construction — it is observed that the large firm may be moving toward a position of permanent dominance.[9]

It would be desirable, therefore, to include some programming examples portraying the effects of economies of scale and of diseconomies where they too might exist. This would assist not only in

7. The median firm in the size-class producing 10–24 houses per year earned $18,400 return to the entrepreneur. Cf. Maisel, *op. cit.*, p. 38.

8. Chamberlin, *op. cit.*, 6th ed., Appendix B, *passim*, for general discussion of this point. A recent empirical investigation by Joe S. Bain, "Economies of Scale, Concentration, and the Condition of Entry in Twenty Manufacturing Industries," *American Economic Review*, XLIV (1954), 15–39, brings forward estimates of large minimum plant-size for economically efficient operation in many industries.

9. Maisel, *op. cit.*, produces evidence (pp. 189–213) to show how builders' unit costs would probably vary downward with increases in the number of identical houses built of a single design. The Bureau of Labor Statistics estimated that the 1949 output of large builders (i.e., builders of one hundred or more houses annually) constituted about 40 per cent of all dwelling units built in U. S. metropolitan areas for the private market in that year, after a correction was made for non-professional building.

Data from Housing and Home Finance Agency — BLS, "120,000 Firms in Residential Building Industry in 1949" (News Release dated August 25, 1951, Washington, D.C.).

the examination of the immediate problem but in drawing inferences from this paper about the more general applicability of the present analysis.

The fact is, however, that abandoning the assumption of constant returns to scale makes for profound difficulties in the use of mathematical programming. The present writer understands that some progress is being made in the development of theoretical solutions for dynamic programs, where processes must be hooked together in a series of dated sequences, and in similar extension of the underlying mathematics for the maximizing of nonlinear expressions subject to linear constraints. But it has not seemed possible to intrude upon the *linear* character of the matrix of constraints.

In a very limited way, we can make a place for economies and diseconomies of scale in programming models of the firm. The following case rests on hypothetical activity-coefficients, and the solution method is indicative of the difficulties. We assume, as before, two limiting resources, executive organization and working capital, and a single type of physical output. We have the following equation system:

$$R = 3\,x_1 + 2\,x_2 + 3\,x_3 + 0\,x_4 + 0\,x_5 \tag{1}$$

is to be maximized subject to

$$10\,x_1 + 4\,x_2 + 8\,x_3 + 0\,x_4 + x_5 = 100 \tag{2}$$
$$5\,x_1 + 7\,x_2 + 5\,x_3 + x_4 + 0\,x_5 = 100 \tag{3}$$
$$x_i \geqslant 0,\ (i = 1, \ldots\ldots, 5), \tag{4}$$

and
$$x_3 \geqslant 10,\ \text{if } x_3 \neq 0. \tag{5}$$

The existence of an economy of scale is indicated by two facts. The third activity, whose level of operation is indicated by the value of x_3, is unquestionably "better" than the first activity, since the former's resource coefficient is the lower in the equation (2) while in equation (3), which represents the burdens on the second constraining resource, the two coefficients are equal. Equation (5), however, says that the third activity must be operated either at zero-level or at or above a critical minimum of ten units. Thus we have built into a programming problem a rough approximation to increasing returns to scale involving a discontinuous shift to lower resource requirements at the critical minimum activity-level.

The method of solution illustrates the limitations of programming mathematics. First, solve for the optimal program in the usual way without, however, including any effect from equation (5). This optimal solution may not violate the special constraint, in which case

we have no interesting difficulties. In the problem given, the optimal solution, without reference to (5), is ($x_2 = 8.33$; $x_3 = 8.33$), which does violate the constraint. We now split the problem into two parts: (a) find the optimal program for $x_3 = 0$; and (b) find the optimal program for $x_3 \geqslant 10$. Comparing the total net revenues of (a) and (b), we choose the larger. In the present illustration, the solution of part (a) is ($x_1 = 6$; $x_2 = 10$; $R = 38$); the solution of part (b) is ($x_2 = 5$, $x_3 = 10$, $x_4 = 15$, $R = 40$). Thus, the preferred solution does require that the third activity operate at its critical minimum.

Discouragement is accorded by the fact that the number of separate *programs* to be solved is doubled as we admit just one possible pair of alternatives involving economies of scale. The approach does not appear promising as a way of incorporating any significant number of such possibilities into programming models.

IV. CONCLUSIONS

We have now completed this survey of applications of linear programming techniques to the scale-integration problem in residential construction firms. A summary of the major points is in order.

1. The firm's own working capital fund and its executive group can be considered the prime resource limits which operate to produce a joint optimum involving both the correct scale of operation and the appropriate internal organization of the firm.

2. The linear programming technique is suited to the task of approximating the conscious decision-pattern of entrepreneurs.

3. Extensions can be made of this linear programming model to cover other important problems of choice: the price-classes toward which output should be directed; the appropriate distribution of the use of machinery and other capital equipment; the optimal distribution of the efforts of salaried employees or other associates of the entrepreneur in an augmented central organization; and the best application of the entrepreneur's own time in a refined breakdown of the functions he performs.

4. Sets of approximate empirical coefficients for the use of capital and of executive time were presented; by solving various programs for different combinations of these coefficients, we have shown what specific alterations in the scale of output and the internal organization of the firm occur when capital funds are increased. This analysis is useful in depicting the growth pattern that may be achieved by the firm during a succession of production periods if R, *ex ante* net revenue, is actually realized, and if a definite proportion of it is then retained in the firm. Two or more firms having different sizes of capital fund

may also be compared, at their respective optima, as to size of output and type of internal organization.

5. The central significance of the turnover rate on working capital has been demonstrated. The same optimal program can be supported by widely different sizes of capital fund, depending on the size of the turnover rate and the policy of the firm with respect to cash withdrawals, and payments of salary to permanent personnel, during the production period. The importance of a tight rein on cash withdrawals, especially for the undercapitalized firm, is clearly shown; and it is evident that, where *ex ante* net revenues approximate *ex post* results, the firm with a tight policy can grow much more rapidly than can the firm whose entrepreneur must live well in the short run.

6. The augmented central organization has been shown to have a quite different internal distribution of functions at the executive level than does the one-man firm having the same amount of capital funds. It is believed that extensions of the present technique to more complicated organizational situations would be fruitful.

F. E. BALDERSTON.

UNIVERSITY OF CALIFORNIA,
BERKELEY

THE STERLING AREA'S CENTRAL POOLING SYSTEM RE-EXAMINED

By Elliot Zupnick

Introduction, 71. — I. The sterling area mechanism, 71. — II. The central pooling system and inflation, 75. — III. The central pooling system and the misallocation of sterling area resources, 82. — IV. Conclusions, 84.

The central pooling of hard currency earnings which characterized the postwar sterling area arrangements has attracted considerable attention from students of international finance. Much of the literature concerned with this aspect of sterling area machinery, however, has been limited to an analysis of two problems: The relative advantage of the pooling system to the United Kingdom and to the rest of the sterling area; and the discriminatory nature of the arrangements. Very little attention has been devoted to an analysis of two major defects of the system as it operated in the postwar period: The tendency of the central pooling system to contribute to inflationary pressures in the sterling area; and the manner in which the system contributed to the misallocation of scarce resources regarded from the vantage point of the area as a unit. The object of this paper is to examine these two aspects of the central pooling system and to indicate their significance for current sterling area policy. Before this is done, however, it is necessary to review briefly some salient features of the postwar sterling area mechanism.

I. THE STERLING AREA MECHANISM[1]

The postwar sterling area was an outgrowth of wartime developments and differed in many fundamental respects from the prewar sterling bloc. Members of the sterling bloc[2] for one reason or another[3]

1. The reader is referred to K. M. Wright's article, "Dollar Pooling in the Sterling Area, 1939–1952," *American Economic Review*, Sept. 1954, for a detailed analysis of the technical aspects of the sterling area's dollar pool.
2. The sterling bloc, unlike the sterling area, was never formally defined. It is generally agreed, however, that the bloc consisted of the British Commonwealth exclusive of Canada, the Scandinavian countries, Egypt, the Sudan and Iraq, the Baltic countries, Hungary, Greece, Yugoslavia and Portugal. Japan and Argentina, though not considered members of the sterling bloc, tended to fix their official rates in terms of sterling.
3. See the League of Nations, *International Currency Experience*, 1944, pp. 48 ff.

found it advantageous to hold their reserves in the form of sterling, to maintain their currencies in a relatively fixed relationship with the pound and to use London's banking facilities to discharge obligations incurred on international account.   One of the characteristic features of the sterling bloc was that it was a voluntary association of countries virtually devoid of control from the center.   There were no formal or informal "rules of the game" to which member countries were legally or morally bound to adhere.   Sterling countries were not required, for example, to hold any specified proportion of their reserves in London, nor were they obliged to maintain their currencies in a fixed relationship with the pound.   The relative stability of exchange rates that did in fact prevail within the sterling bloc between 1931 when Britain abandoned the gold standard and the outbreak of the war in 1939 is attributable to the pursuit of self-interest by the associated countries; the United Kingdom did nothing to maintain this stability.[4] Finally, membership in the sterling bloc resulted from unilateral action and neither conferred special privileges nor entailed obligations.   Each participating country maintained full sovereignty over its currency and the United Kingdom exercised neither direct nor indirect control over the economic, financial or commercial policies of the associated members.

The outbreak of the war in 1939 resulted in three major changes in the sterling area mechanism.   First, the composition of the area was drastically modified as non-Commonwealth countries with the exception of Egypt, the Sudan, Iraq, Iceland and the Faroe Islands severed connections with sterling.   Secondly, the area obtained formal status as a result of the exchange control regulations which were promulgated shortly after the outbreak of the war.   Finally, the area was transformed into an exchange control system.   This modification was by far the most important and was largely responsible for the sharp difference between the prewar sterling bloc and the wartime and postwar sterling area.

The major function of the wartime sterling area was to aid in the mobilization of financial and economic resources in general and in the conservation of hard currency exchange in particular.   The sterling area countries attempted to achieve these objectives in three ways: first, the United Kingdom and the rest of the sterling area issued exchange control regulations designed to prevent unauthorized monetary flows from sterling to nonsterling countries while maintaining the maximum degree of freedom within the area; secondly, sterling

4. See W. A. Brown, *The International Gold Standard Reinterpreted*, (New York: National Bureau of Economic Research, 1940), p. 1165.

area countries tacitly agreed to accumulate sterling balances rather than demand immediate payment for goods and services exported to the United Kingdom; finally the hard currency pool was established to conserve scarce reserves and to ensure their utilization in a manner best calculated to serve the war effort.

The establishment of the hard currency pool represented a sharp break with prewar practices. For though members of the sterling bloc customarily converted their excess foreign exchange earnings and newly mined gold into sterling, there was no obligation to do so. Furthermore, the convenience of holding international reserves in the form of sterling did not imply discriminatory treatment in favor of sterling goods, and the full convertibility of sterling at the going market price assured that this result would not eventuate. Finally, sterling bloc countries were guided by the state of their individual sterling reserves in framing their economic, financial and commercial policies and not by the state of Britain's gold reserves which was important to them only to the extent that it reflected the soundness of Britain's position as an international banker. The contrary was true after the establishment of the central pool: Sterling area countries were obliged to sell their hard currency earnings to London; discrimination against imports from hard currency sources and in favor of sterling goods was implicit in the operation of the central pool;[5] and sterling area countries were expected to frame their relevant economic policies in the light of the state of the central reserves rather than with reference to the state of their individual sterling reserves. The significance of this latter feature of the wartime and postwar sterling area arrangements is discussed in detail below.

With respect to the right to draw upon the central pool's resources, sterling area policy underwent some changes during the postwar period. During the war the associated countries enjoyed full convertibility of sterling. The task of restraining the demand for hard currencies was assumed by the *local* trade authorities rather than by the exchange control. This is to say that, provided the trade authorities issued a license approving a transaction involving the payment of hard currencies, resident account sterling was freely convertible.

The technically full convertibility of resident account sterling which prevailed during the war was limited for a brief span of time in the postwar period. Immediately after the cessation of hostilities, the United Kingdom negotiated agreements with Egypt and Iraq

---

5. The American authorities' opposition to the discriminatory aspects of the central pooling system led to the provision in the Anglo-American Financial Agreement which called for its abolition.

specifying the amount of hard currencies they could withdraw from the central pool. The heavy drain of dollars during the brief convertibility period in 1947 led to an extension of this practice. Britain ultimately negotiated hard currency allocation agreements with India, Pakistan, Burma and the Irish Republic. Of the sterling area Dominions, Australia and New Zealand alone enjoyed full convertibility throughout the postwar period. In view of the fact that the economic, financial and commercial policies of the Dependencies were framed in London, it was not necessary to limit their convertibility rights by agreement.

The limitations imposed upon the rights of the newer Dominions to convert their sterling into hard currencies was short-lived; by the end of 1950 the practice of negotiating hard currency agreements was abandoned and technical full convertibility of resident account sterling was restored. Several factors were responsible for this reversal of policy. First, the agreements were honored more in their breach than in their observance; India and Pakistan, for example, consistently drew more than was formally agreed upon. Secondly, the practice of discriminating against the newer Dominions with respect to their rights to draw upon the resources of the central pool created thorny problems and was not designed to foster the harmony that was necessary for the successful operation of an exchange control system. Thirdly, this method of staying the drain on the reserves was not particularly successful and hence did not gain many adherents in sterling area circles. Finally, India and Pakistan, two of the most important sterling area countries subject to the allocation accords agreed in July 1949 to reduce their dollar imports to 75 per cent of the previous year's level[6] with the result that the allocation accords lost much of their *raison d'etre*. In any event, by the end of 1950, resident account sterling was again freely convertible within the limits established by the local trade authorities.

The freedom to draw upon the resources of the central pool which outer sterling area countries enjoyed in the postwar period impeded the attainment of the objectives which motivated the establishment of the pooling system. Furthermore, the absence of effective co-ordinating machinery and the high degree of decentralization which characterized the postwar sterling area arrangements precluded effectual control from the center and compounded the area's problems. The remainder of this paper is concerned with an analysis of two major ways in which the defective sterling area machinery

6. The calendar year 1948 was the base period for India while the year ending in June 1949 was the base period for Pakistan.

accentuated the area's economic difficulties. Attention is focused first on the manner in which the operation of the central pooling system contributed to the inflationary pressures in the sterling area and then on the manner in which it eventuated in a misallocation of resources.

II. The Central Pooling System and Inflation

The practice of pooling hard currency earnings and financing deficits through drawings on the resources of the central pool contributed to the inflationary pressures in the overseas sterling area countries by effecting a partial divorcement between the financial, economic and commercial policies pursued by these countries and the *direct* responsibility for the financing of the possibly ensuing deficits on international account. The partial substitution of group for individual responsibility implicit in the operation of a central pooling system would not by itself be unfeasible if the associated countries were bound together by a strong community of interest, or if the economic, financial and commercial policies of the affiliated countries were framed with reference to similar objective criteria, or if effective control from the center precluded the possibility of some countries pursuing policies which were at variance with the group's welfare. None of these conditions prevailed in the postwar period: An examination of the economic aspirations of the sterling area countries reveals important areas of conflict; the state of the central reserves exercised a varying influence on the policies of the associated countries; and the sterling area retained to a large degree the decentralized features of the prewar sterling bloc.

Ordinarily, one of the most effective restraints on a country living beyond its means (measured in terms of its productive capacity and its ability to obtain autonomous loans) is the ensuing international deficits which must be financed, in the long run, by a reduction of its reserves. Provided a country is not "working off" excess reserves, the necessity to stay the drain on the reserves must lead to the introduction of contractionary policies designed to reduce the level of domestic expenditure to a point where it can be satisfied out of the resources made available by domestic production and autonomous loans.[7] It is of little significance in this context whether the reduction in demand is effected by the decline in reserves operating directly on the money supply or indirectly through its influence on the rate of interest or whether it results from the introduction of dis-

7. The term autonomous loans is used to distinguish this type of borrowing from induced short-term loans which are equilibrating in character.

inflationary budgetary policies. In any event, the level of domestic expenditure must be reduced. And it is this necessity which precludes an open economy from pursuing more inflationary policies than its neighbors.

The central pooling of hard currencies initiated by the sterling area countries during the war and projected into the postwar period seriously weakened the restraints normally imposed by balance-of-payments disequilibria. For under this system, the state of the central reserves rather than that of the individual members had to play the decisive role in the formulation of the relevant economic and commercial policies. There were two factors in the postwar period

TABLE I

STERLING RESERVES OF THE OVERSEAS INDEPENDENT STERLING
AREA COUNTRIES

(Million U.S. Dollars)

	1937	1946	1948	1950	1952
Australia............	291	677	1126	1432	789
New Zealand.........	68	n.a.	213	143	150
Burma.............	—	—	113	119	n.a.
Ceylon.............	—	—	254	191	163
Ireland.............	57	170	192	228	202
Iraq...............	26	179	139	117	129
India[1]..............	316	4905	3099	1752	1482
Pakistan...........	—	—	787	479	n.a.

Source: Derived from International Monetary Fund, International Financial Statistics.
[1] Prior to 1948, India's reserves include those of Burma, Ceylon and Pakistan.

which militated against this arrangement. The first was that while the central reserves were almost constantly at precariously low levels dictating policies of restraint, the sterling reserves of the self-governing outer sterling area countries were extraordinarily large and imparted an inflationary bias into the system (see Table I). The second factor which precluded the successful operation of a central pooling system in the postwar period was the determination on the part of the overseas independent sterling area countries to diversify their economies almost at any cost in order to render them less susceptible to vicissitudes over which they have no control and which were largely regarded as being responsible for their failure to share equitably in the fruits of economic progress in the interwar period. This strong desire for diversification through industrialization re-enforced the inflationary bias imparted to the system by the huge sterling reserves of the associated countries by making them more prone to frame their economic, financial and commercial policies in

the light of the state of their own reserves rather than with reference to the state of the central reserves.

The importance of the accelerated pace of development in the outer sterling area as a contributory factor to the dollar drain is suggested by the data presented in Table II which shows the increase in the volume of select capital good imports that the overseas independent members of the sterling area derived from the United States between the prewar and postwar periods. Examination of the data in Table III reveals, moreover, that the increase in the volume of capital good imports derived from the United States was not due to the inability of the outer sterling area to obtain these imports from traditional sources. On the contrary, capital good imports from the United Kingdom, which in the prewar period was the major supplier, increased spectacularly between 1938 and the postwar period.

It should be noted that it was not the pooling system, *per se*, that contributed to the sterling area's difficulties but rather the economic environment in which the system operated and the lack of

TABLE II

OUTER STERLING AREA COUNTRIES' IMPORTS OF SELECT CAPITAL GOODS
FROM THE UNITED STATES

(Million U.S. Dollars in 1948 Prices)

	Metals and Manufactures	Machinery	Transport Equipment[1]	Chemicals
Australia:				
1938.............	9.9	28.0	8.4	5.2
1948.............	12.9	25.8	16.2	5.6
1951.............	19.0	35.5	30.1	8.9
New Zealand:				
1938.............	2.0	12.8	2.1	1.2
1948.............	3.8	11.2	1.8	1.9
1951.............	4.1	9.9	9.2	3.1
India:[2]				
1938.............	6.5	13.6	4.6	6.3
1948.............	35.2	63.3	38.0	38.2
1951.............	15.1	41.4	27.8	26.0
Iraq, Burma and Ceylon:				
1938.............	1.7	3.2	.7	.8
1948.............	1.6	7.9	3.7	1.7
1951.............	16.3	5.1	3.2	3.5

Source: United Nations, Economic Commission for Europe, *Economic Survey of Europe Since the War* (Geneva, 1953), Table LV, p. 313.
[1] Exclusive of personal cars.
[2] Including Pakistan.

TABLE III

OUTER STERLING AREA COUNTRIES' IMPORTS OF SELECT CAPITAL GOODS
FROM THE UNITED KINGDOM
(Million U.S. Dollars in 1948 Prices)

	Metals and Manufactures	Machinery	Transport Equipment[1]	Chemicals
Australia:				
1938.	55.1	65.6	42.5	20.2
1948.	52.2	108.4	63.4	20.7
1951.	104.9	199.2	179.2	41.3
New Zealand:				
1938.	30.9	26.0	15.1	8.0
1948.	30.8	44.1	21.9	7.5
1951.	42.7	53.7	43.0	14.9
India:[2]				
1938.	43.0	85.2	20.7	27.6
1948.	50.0	174.8	43.7	42.7
1951.	60.6	180.5	81.7	58.8
Iraq, Burma and Ceylon:				
1938.	15.3	13.1	7.6	5.1
1948.	28.3	37.5	22.2	10.9
1951.	22.1	27.9	19.3	16.8

Source: United Nations, Economic Commission for Europe, *Economic Survey of Europe Since the War, op. cit.,* Table LV, p. 308.
[1] Exclusive of personal cars.
[2] Including Pakistan.

effective safeguards which were required to ensure success in the absence of ideal conditions. A fundamental disharmony existed in the postwar period between the individual associated countries and the group as a whole which introduced an important element of instability and seriously impaired the possibility of achieving the objectives which motivated the formation of the central pool. It was noted above that the pool was established to aid in the conservation of scarce resources. The underlying idea was that a combination of reserves would render the area both stronger and more stable than the sum of its component parts. A high degree of discipline was required for the attainment of these objectives. This discipline was forthcoming during the war for a variety of reasons, chief among which were the desire to win the war. In addition, however, the severe shipping shortage effectively restrained the demand for imports by the outer sterling area countries. Under these circumstances, the pooling of reserves contributed significantly to the economic strength of the area. The disappearance of these restraints in the postwar

period in conjunction with the strong urge to industrialize and the swollen sterling reserves which justified the activation of this urge militated against the attainment of the necessary degree of discipline.

One of the anomalies of the postwar sterling area arrangements was the failure to perceive that in the absence of a high degree of discipline to ensure that individual sterling countries would not pursue policies inimical to the group's welfare, some control from the center was essential. In the event, this control was not forthcoming; the voluntary and individualistic aspects of the prewar sterling bloc were established as leading tenets of the postwar sterling area mechanism.

It is ironic to note that the Organization for European Economic Cooperation (O.E.E.C.), formed to allocate Marshall aid among the recipient countries, succeeded in establishing a higher degree of co-ordination among its members than did the postwar sterling area. As a basis for deciding how to allocate the global quota voted by the American Congress, the O.E.E.C. required each participating country to submit its economic program for the forthcoming year. These programs were carefully scrutinized and within broad limits helped determine the allocation of aid.

In the early postwar period, there was nothing in the sterling area mechanism which approximated even the crude co-ordinating machinery of the O.E.E.C. Indeed, until late 1952, sterling area machinery consisted of little more than a statistical committee composed of technical representatives of the associated countries[8] which was formed to disseminate relevant balance-of-payments information, the Commonwealth Liason Committee which had no policy-making function but which served as a "forum where policy matters can be discussed and general approval secured for decisions,"[9] and *ad hoc* conferences of high-level sterling area officials which were convened at irregular intervals to map emergency action to cope with specific problems as they arose.[1] It is manifest that this machinery was inadequate for the formulation of long-run plans and the co-ordination of policies that were required in the face of the persistent dollar problem and the inflationary bias imparted to the system by the

---

8. Canada, not a member of the sterling area, also had a representative on the statistical committee.

9. A. R. Conan, *The Sterling Area* (London: Macmillan and Co. 1952), p. 159.

1. The first high-level conference was convened immediately after the suspension of convertibility. Other conferences were held: in July 1949 to adopt concerted action in order to stay the drain on the reserves, in September 1950 to allow for a relaxation of import restrictions in the light of favorable reserve developments, and in January 1952 to deal with the crisis which resulted from the collapse of raw material prices.

swollen sterling reserves and the powerful and all pervasive desire to industrialize.

Partially as a consequence of the recognition of the need for improved co-ordinating machinery, Churchill called a conference of Commonwealth finance and prime ministers to meet in December 1952 "to review together the pressing issues of financial, commercial and economic policies with which our several governments are faced."[2] This represented the first time in the postwar period that a Commonwealth conference was convened to discuss the broader problems facing the sterling area. To the extent that one can judge from the communique published at the conclusion of the discussions, the conference was primarily concerned with the problems of sterling area development and with the nature of the policies required to allow for an ultimate restoration of sterling convertibility.[3] These also appear to have been the major topics discussed in a subsequent conference held in Sydney, Australia in January 1954. Though the communiques issued at the conclusion of both of these conferences stress the need for a co-ordinated sterling area development program, there is no indication that the central pooling arrangements were discussed or that any modifications of the sterling area mechanism that would make it more consistent with the attainment of the explicitly stated objectives were entertained.

In the absence of the necessary degree of discipline, of co-ordinating machinery or of central control, the overseas independent sterling area countries were heavy drawers on the resources of the central pool. Between 1946 and June 1953 these countries drew 1,285 million dollars from the central pool.[4] This figure is the net resultant of the direct dollar deficits of the self-governing outer sterling area countries and their gold sales to the United Kingdom. It does not include, however, any part of the dollar and gold payments the United Kingdom made on their account to the non-dollar countries. Were this data available, the drain on the central reserves attributable to the overseas sterling area countries would be even larger.

It would be erroneous to conclude that the drawings of the outer sterling area countries on the central pool measure the full extent of the drain on sterling area resources attributable to the operation of

2. British Information Service, *Labor and Industry in Britain*, Sept. 1952, p. 3.

3. Cmd. 8717, Commonwealth Economic Conference: Final Communique, Dec. 1952.

4. Derived from United Kingdom Balance of Payments, 1946–1953, Cmd. 8976 H.M.S.O. London, Oct. 1953, Table 20, p. 46.

the pooling system. This is due to the fact that the volume of domestic expenditure which the availability of a specified amount of hard currency reserves can support is a multiple of these reserves; the multiplier is determined by the marginal propensity to import hard currency goods. Thus if 10 per cent of an increment to income is devoted to the purchase of dollar goods, the availability of x dollars would enable an expansion of domestic expenditure by $10\,x$. The ultimate drain on sterling area resources is thus measured not by the direct drawings on the central pool alone, but by these drawings in conjunction with the complementary resources obtained either from the United Kingdom or from domestic sources. To the extent that the resources absorbed by the self-governing countries of the outer sterling area as a result of their ability to draw on the central pool precluded the expansion of their own or of Britain's hard currency exports, this drain was no less significant than that represented by the direct drawings on the central reserves.

An appreciation of the indirect burden imposed by the operation of the pooling system focuses attention on a possible relationship that is not readily apparent: namely, that the huge volume of unrequited exports the United Kingdom sent to the self-governing outer sterling area countries as a result of their drawing down their war-accumulated sterling balances may not have been unrelated to the relatively free access to the central pool which these countries enjoyed. Had these countries not been able to draw so freely on the resources of the central pool, their demand for complementary goods from the United Kingdom might have been considerably reduced. The possibility of this occurring is re-enforced by the consideration that one of the rules of the postwar sterling area dictated that imports were not to be derived from hard currency sources if alternative goods were available in sterling countries. Thus the inability to obtain imports from hard currency sources might have effectively choked off the demand for these goods and inhibited those projects contingent upon them.

Recognition of the indirect burden imposed by the operation of the central pooling system also throws some doubt on the validity of one of the justifications Britain advanced for the rather heavy release of sterling balances to the self-governing countries in the post-war period: namely, that these releases did not contribute to the area's economic difficulties but were, rather, dollar savers.[5] It may be, however, that this argument represents a case of putting the cart before the horse. In the light of the argument advanced above, it is possible that the heavy drawings which these countries were per-

5. Hansard, V. 448, c. 207.

mitted to make on the central pool's resources were a factor respon-
sible for the large volume of unrequited exports and that these exports
were not dollar savers but were themselves induced by the ready
availability of hard currencies.

If these considerations are valid, it would appear that the freedom
with which the overseas sterling area countries drew upon the central
pool's resources not only accentuated their own inflationary pressures
by enabling them to initiate projects which in the absence of hard
currency reserves they could not have undertaken, but also intensified
the inflationary pressures in the United Kingdom by inducing a
larger volume of unrequited exports than would otherwise have
occurred. A more restrictive policy might thus have enabled a
general expansion of sterling area exports to the hard currency
countries.

III. The Central System and the Misallocation of Sterling Area Resources

In addition to allowing inflationary programming in the asso-
ciated countries, the central pooling system contributed to the mis-
allocation of resources regarded from the vantage point of the area
as a unit. Three major factors were responsible for this development:
first, the lack of effective co-ordinating machinery frequently resulted
in the countries pursuing the most inflationary programs obtaining a
disproportionate share of the available hard currency; secondly, the
fact that the state of the central reserves was more effective in
restraining some sterling area countries than it was with respect to
others accentuated this tendency; and finally, the fact that no sterling
area country was able to ignore the state of the central reserves during
the crises periods frequently resulted in the abandonment of projects
before they were brought to fruition and thus led not only to a mis-
allocation of resources but also to their waste.

It is not necessary to discuss in detail the manner in which the
first factor cited above contributed to the misallocation of resources
regarded from the vantage point of the area as a unit. It was already
noted that except for a brief period of time, the amount of hard cur-
rency any particular sterling area country was able to draw from the
central pool was determined by its own trade authorities. One
important consequence of this arrangement was that the countries
that were pursuing with greatest abandon their development pro-
grams and in which the restraints against inflation were least effective
were the largest recipients of capital in the form of drawings from the
central pool. Since these drawings were in the nature of an ex-post

flow induced to finance deficits already incurred, it would have been entirely fortuitous had the actual distribution of capital coincided with that which would have resulted from a co-ordinated and sound development program embracing the entire area. There is no reason to believe that this happy result was actually achieved.

The misallocation of resources that inevitably resulted from the failure to co-ordinate sterling area development programs was intensified, moreover, by the fact that the state of the central reserves exercised a more restraining influence on some components of the area than on others. The economic, financial and commercial policies of the United Kingdom and the Dependencies, for example, were intimately affected by the state of the central reserves. In the former case this is attributable to the fact that there is no dichotomy between Britain's reserves and the central reserves; the United Kingdom's reserves consisted of the gold and dollar holdings of the Exchange Equalization Account. In the case of the Dependencies, though a dichotomy existed between the central reserves and their individual sterling reserves, this was not a source of difficulties for the sterling area; the economic, financial and commercial policies of the Dependencies were framed in London by officials who took the cue from the state of the central reserves.

In contrast to the situation in the United Kingdom and the Dependencies, the influence of the state of the central reserves on the relevant policies of the self-governing countries of the outer sterling area was considerably less strong. As was noted above, the existence of huge sterling reserves imparted an inflationary bias into the system by making these countries more prone to frame their policies in the light of their own reserves rather than with reference to the state of the central reserves. This varying significance of the state of the central reserves as a determinant of policy implied that under certain circumstances some capital projects in the United Kingdom and the Dependencies would have to be sacrificed without regard to their importance from the vantage point of sterling area development while other projects, perhaps of considerable less usefulness, were continued in the independent sterling area countries. A recognition of this possibility prompted the *Economist* to comment that Australia's ambitious investment program was partially financed by the natives of the Gold Coast.[6] It is interesting to note in this context that the Dependencies contributed 2,155 million dollars to the central pool between 1946 and June 1953.[7]

6. *The Economist*, Oct. 11, 1952, p. 114.
7. Derived from *United Kingdom Balance of Payments, 1946–1953*, Cmd. 8976, *op. cit.*, p. 46.

The third major way in which the operation of the central pooling system contributed to the misallocation of resources is related to the fact that even the self-governing countries were unable to ignore the precariously low levels of the reserves during the crisis periods. The major characteristic of each of the postwar crisis years was a precipitous drain on the reserves which had to be stayed if catastrophe were to be avoided. The sterling area responded to these crises by hastily convening high-level conferences to initiate emergency measures. In the event, these measures invariably consisted of the intensification of import restrictions. While this method of staying the drain succeeded with respect to the immediate objective, it did so only at the expense of failing to carry through to completion some projects that were already initiated. The misallocation or waste of resources that resulted from the abandonment of projects could have been avoided had there been a proper degree of planning and co-ordination of sterling area development programs to ensure that projects once commenced would obtain the necessary finance to bring them to fruition. The substitution of emergency action for long-term planning and co-ordination of even the most elementary nature was thus an important contributory factor to the misallocation and waste of scarce resources.

IV. Conclusions

An examination of the sterling area's postwar central pooling arrangements has revealed serious defects which contributed both to the intensification of inflationary pressures and to the misallocation of resources. These findings have an important bearing on current sterling area policy as promulgated in the communiques issued after the last two Commonwealth economic conferences. This policy places heavy emphasis on the need for a co-ordinated sterling area development program designed to reduce and ultimately to eliminate the area's balance-of-payments deficits. Toward this end, the United Kingdom pledged to increase the flow of capital to sterling area countries to help finance projects which can "contribute to the improvement of the area's balance of payments difficulties." The self-governing territories for their part recognized the evils of inflation and expressed their determination "to persevere in their efforts to curb [it]." Though these statements of intent are admirable, it is feared that their execution may be frustrated unless the central pooling system is modified so as to reduce the inflationary bias inherent in the sterling area mechanism and to prevent its operation from resulting in a misallocation of sterling area resources.

<div align="right">Elliot Zupnick.</div>

City College of New York

HOUSING: HAS THERE BEEN A DOWNWARD SHIFT IN CONSUMERS' PREFERENCES?

By Louis Winnick*

I. Residential Housing Since 1890

In any economy characterized by substantial long-term increases in the standard of living the production of necessities does not generally keep pace with the total output of goods and services. Consumer expenditure patterns are altered as new products emerge and luxury goods become more widely distributed. Kuznets has shown that the output of goods which can be classified as "old" and "indispensable," taken as a proportion of total consumer goods, has declined, at least since the latter part of the nineteenth century, while the share of "new" and "dispensable" goods has risen.[1]

Similarly, residential construction outlays (including expenditures for additions and alterations but not repairs and maintenance) measured as a proportion of either gross national product or gross capital formation have experienced a marked decline since about 1890 in addition to the more familiar major swings. Even the current housing boom has not attained the relative strength of the 1920's and falls considerably short of the relative levels achieved during the 1890's. Since residential capital formation is closely geared to increases in population and households, the declining rate of population growth in itself would provide a sufficient explanation for its diminishing relative importance without further examination of the role of the consumer.

But the burden of this paper is that the decline of nonfarm housing construction in total output, rendered inevitable by demographic trends, has been powerfully reinforced by changes in consumer behavior. For not only has housebuilding declined in relation to total

1. Simon Kuznets, *Longterm Changes in the National Product of the United States*, International Association for Research in Income and Wealth, Part V (1951), p. 33.

production, but the per capita real value of the standing stock of housing has failed to show any marked trend during the first fifty year of this century, and the average real value per dwelling unit standing has declined perceptibly (Table I).   These phenomena suggest, at least, that there has been a downward shift in consumer preferences for housing.

TABLE I

VALUE PER CAPITA AND PER DWELLING UNIT OF RESIDENTIAL CAPITAL, 1890 TO 1950 (1929 DOLLARS)

Year	Per Capita Value	Per Dwelling Unit Value
1890	$658	$2,651
1900	793	3,355
1910	792	3,320
1920	747	3,119
1930	870	3,135
1940	770	2,662
1950	740[1]	2,381[1]

Source: Grebler, Blank and Winnick, *op. cit.*, and Table IV.   A summary explanation of the derivation of these estimates is given in the Appendix.
1. The maximum estimates, based on the wealth total derived from the 1950 Census of Housing would be $775 and $2,491. These estimates undoubtedly err on the high side. At any rate the main line of the argument would remain unchanged if the higher values were substituted.

A moderate rise in per capita value took place between 1890 and 1910 followed by a decline during the World War I decade, a sharp rise resulting from the housing boom of the twenties, and a pronounced drop since 1930. The per capita value of housing in 1950 falls considerably below the 1900 value. Though it may be astonishing to find the 1950 value below 1940 in spite of an impressive postwar residential construction boom, it should be recalled that new construction was severely limited during the war years and that a very large proportion of new households even in the years following 1945 was accommodated by conversion of existing houses; conversions represent an extremely economical method of adding to the stock of dwelling units. It is probable that there has been some understatement in estimates of construction expenditure for the decade of the forties which, together with some minor technical differences in definition and coverage, lead to some understatement in the 1950 values. It is unlikely, however, that any reasonable adjustment for these two factors would cause the 1950 per capita value to approximate the 1900 level, and thus leave the inference that we have not increased our use of housing resources over the past half-century.

As Table I shows, real value per dwelling unit has performed

even more poorly than per capita value, primarily because of the long-run decline in household size which permits a smaller dwelling unit. The decline in average real value per dwelling unit standing reflects in part the fall in the average real input of *newly constructed* dwelling units. Average real input declined approximately 40 per cent during the past six decades.[2]

In interpreting both per capita and per dwelling unit figures, one must bear in mind that the average age of the housing inventory has been slowly increasing. The annual depreciation charge (described in the Appendix) increases proportionately with the stock of housing but value increments have been declining relative to stock. The result is that capital consumption has become a very large and increasing offset to gross residential capital formation.

The findings on average real values per capita and per dwelling unit are astonishing in view of the rise in per capita real income over the past fifty years and (1) the consumption trends for other necessities and (2) the income and price elasticities of housing. While the shares of other necessities have diminished, rising standards of living have called for increased per capita expenditures on them. Thus, the share of perishables (of which food is a major component) has declined but per capita outlays have increased substantially from $127 (in 1929 prices) in 1884–1893 to $233 in 1929–1938.[3]

Second, it is a well-known fact and observable in cross-sectional data, that capital expenditures on housing have positive income elasticity. That is, the average value of a purchased home rises with the income of the purchaser though at a slower rate. This relationship is illustrated by FHA data in Table II. Because per capita real income has more than doubled since 1890, it would be entirely reasonable to expect some discernible rise in the per capita use of housing capital.

Table II offers a clue to the size of the income elasticity coefficient. It may be noted that with a roughly doubled income from $1,000–$1,499 to $2,000–$2,499, from $1,500–$1,999 to $3,000–$3,499, and from $2,500–$2,999 to $5,000–$6,999 the percentage value increase of existing homes ranges from 42 to 67 per cent. From data in the 1950 Housing Census[4] coefficients ranging between 38 and 60 per cent

2. Because this finding came as a surprise, the construction cost index used for all deflated series was tested for a productivity bias. Although attempts to measure errors in cost indexes due to productivity can never be conclusive because of the complexity of the problem, the available evidence brought to bear indicates no serious long-term bias.

3. Simon Kuznets, *National Product Since 1869*, p. 106, Table II-8 and p. 107, Table II-9, col. 1.

4. 1950 Census of Housing, Vol. IV, *Residential Financing*, Part I, p. 272.

TABLE II

RELATIONSHIP BETWEEN AVERAGE PROPERTY VALUATION AND BORROWERS'
ANNUAL INCOME, NEW AND EXISTING SINGLE-FAMILY OWNER-OCCUPIED
HOUSES WITH FHA-INSURED MORTGAGES, 1941[1]

| Income Class | Existing Houses | | New Houses | |
	Average Value[2]	Ratio of Value to Income	Average Value[2]	Ratio of Value to Income
Under $1,000	$ 2,531	312%	$ 2,904	329%
1,000–1,499	3,134	240	3,518	269
1,500–1,999	3,765	216	4,247	244
2,000–2,499	4,463	200	4,976	223
2,500–2,999	5,072	189	5,507	205
3,000–3,499	5,538	177	5,989	191
3,500–3,999	6,208	169	6,506	177
4,000–4,999	7,105	161	7,122	162
5,000–6,999	8,465	148	8,327	148
7,000–9,999	10,398	130	8,890	112
10,000 and over	13,435	98	11,461	82

Source: Federal Housing Administration, *Annual Report*, December 31, 1941, p. 34, reproduced in Ernest M. Fisher, *Urban Real Estate Markets: Characteristics and Financing* (New York: National Bureau of Economic Research, 1951), pp. 84, 87.

1. Similar data are available in FHA annual reports for somewhat earlier periods and for the war and postwar periods. All show the same general picture. Data for 1941 are presented here on the view that this year was characterized by relatively normal relationships in the housing market.

2. Including land. The series would rise somewhat more sharply if value were defined exclusive of land, since other FHA data show that the ratio of land value to total property value rises with the increase in property value.

can be derived, testifying to the stability of the income value relation. In other words, if consumer reactions to income had been similar to that which might have been expected from cross-sectional data, a substantial increase rather than stability in the per capita values might have been expected.

Obviously the deterrent effects of price elasticity cannot be ignored. The rise in both house prices and residential construction costs has been quite striking over the past six decades, substantially higher than in the case of other goods and services. On the basis of nine-year moving averages, the price index implicit in constant and current dollar gross national product series[5] rose by 160 per cent from 1889 to 1945 (central years of the moving average); the residential cost index rose by 340 per cent over the same period, or more than twice as much.

Hardly any data exist on the size of the price elasticity coefficient

5. 1929 = 100. Simon Kuznets, "Nine Year Moving Averages of National Product and Components by Types of Use, 1873–1945," *Capital Requirements Study*, Memorandum No. 19, (New York: National Bureau of Economic Research), July 1951, pp. 4–6.

for housing. Duesenberry and Kistin[6] have estimated the elasticity with respect to *relative prices* for rental expenditures (paid and imputed with real income held constant) to be .078. If this value is applicable, the relative price increase for housing of about 113 per cent may have caused about a 10 per cent shrinkage in housing demand. While more direct data would be highly desirable, it is extremely doubtful that any price coefficient would even approach the income coefficient. If this be so, the presumptive case for a change in tastes would appear quite strong.

Why has the per capita use of housing capital failed to show any marked increase over a half-century? Complete understanding of why the flow of resources into the residential real estate sector has been so insensitive to gains in real income may never be attained. But a few observations are offered.

First, consumers have chosen to spend part of their increments in income not directly on shelter but on many kinds of consumer capital related to the house and considered a vital part of household operation. Statistical measures of residential capital formation are dependent upon an arbitrary definition of items to be included or excluded in measuring outlays on construction, a definition which falls increasingly short of the full outlay for a dwelling unit as seen from the consumer's point of view. Thus, while certain consumer durables, such as oil burners and bathtubs, are captured by the construction data, others, like mechanical refrigerators, washing machines, garden equipment, venetian blinds, etc., are not. An expansion of the definition of residential capital formation to embrace items of this kind might possibly lead to a different conclusion; namely, that per capita use of housing resources has increased.[7]

Thus, while the proliferation of consumer durables has offered serious competition to all other goods in the consumer basket, the

6. James S. Duesenberry and Helen Kistin, "The Role of Demand in the Economic Structure," *Studies in the Structure of the American Economy*, Wassily Leontief, *et al.* (New York: Oxford University Press, 1953).

7. The increase would not seem marked enough, however, to vitiate the inference that some displacement of housing — more broadly defined — has occurred. Per capita consumer capital other than housing increased from $218 (in 1929 prices) in 1900 to $452 in 1948. Raymond W. Goldsmith, "A Perpetual Inventory of National Wealth," *Studies in Income and Wealth*, Vol. XIV (National Bureau of Economic Research, 1951). A large part of the increase in consumer capital is accounted for by automobiles. Because of suburbanization, the automobile has become increasingly associated with home ownership and reduces the "amount" of house a family will undertake to buy. To make a reasonable guess, the average family, if it wished, could afford twice as much house as it now uses if the costs of the automobile (operating and capital) were removed from the family budget.

outlays for these durables — so many of which have become complementary to the dwelling unit — appear to have been substituted for construction expenditures as currently defined in greater measure than for other forms of consumer outlays.

Second, the size of the capital outlay a family is willing to make for a house is not independent of the annual level of carrying costs. To the extent that the latter increase, the former may shrink. If annual costs may be measured by the aggregate rent bill (imputed rents for home owners plus contract rents for tenants), the evidence points to a more rapid rise in rents than in capital values. Rent charges seem to have kept pace with income; that is, in spite of variations from decade to decade, there does not appear to have been any substantial long-run change in the ratio of rent to income; in the meantime the ratio of residential capital to income has been declining. The relatively greater increase in rents is attributed to the proliferation of services other than pure shelter, such as heat, light, janitorial and municipal services, and to the relatively greater amount of depreciation stemming from electric wiring, central heating, and kitchen equipment — all shorter-lived than the shell of a house. A larger and larger proportion of the annual rental bill is occasioned by services other than pure shelter, the costs of which are imputed or actually paid in the form of rent.[8]

Third, the apparent refusal of consumers to allow their rising income to be reflected in larger increases in space per person is also understandable in view of the housekeeping problems created by the large house calling for expensive domestic and maintenance services. At the same time the need for more housing space was becoming less urgent as many activities, recreational and domestic, were shifted from the household as a result of the automobile and the trend towards increased purchase of commodities and services traditionally produced in the home.

Nor need stable per capita values imply that no groups in the community were improving their housing standards in terms of real capital. Many observers feel that the lower income classes have substantially bettered their living accommodation, an observation perfectly compatible with the fact that easy mortgage terms have propelled the relatively poor into the new single-family home ownership

---

8. Since residential rent, the value of the current output of houses, has been rising more rapidly than residential capital, a declining capital-output ratio can be inferred. This is not the place to discuss this measure. It suffices to say that a residential capital-output ratio does not have the same interpretative or predictive value attributed to the capital-output ratios of other sectors because of problems in valuation and the large imputed component of rent.

market more readily than in earlier decades. It is probably among the middle and upper classes that housing has suffered its greatest decline. The ostentatious (and expensive) town house of the 1890's has no modern counterpart. And in any house rent and value distribution one can always find a surprisingly large proportion of families whose housing standards hardly correspond to their income position.

Those who hold the opinion that houses have become increasingly "shoddier" and that the modern house is a poor imitation of the solid structures of yesteryear will be inclined to interpret the data as demonstrating a deterioration of housing quality, at least over the more recent past. Indeed, one could even find a modicum of support in entertaining such a view when it is considered that per capita space increased between 1940 and 1950 while the average value per new dwelling unit decreased, and that rent (which includes repair and maintenance) has been rising relative to capital. But technological advances in material and design interfere with any such interpretation.

It would be erroneous to conclude that consumer satisfaction from housing changes *pari passu* with changes in the quantity of input. A change in fashion which substitutes a fireplace for a foot of ceiling height can raise consumer satisfaction though leading to a net reduction in input. Second, there is no evidence of any decline in the per capita use of space. The size of dwelling units (average number of rooms) has decreased but, apparently, no more rapidly than average household size. While the room is admittedly not the best unit for measuring the quantity of shelter, it is nevertheless significant that the average number of persons per room declined slightly (4 per cent) between 1940 and 1950. Third, even if the per capita use of space (by any reasonable measure) had decreased, how does one balance the effect of greater efficiency in design which may yield more net living space from a smaller gross area by eliminating staircases, useless corridors or dead corners?

Similarly, the lightening of the structural frame of the dwelling lowers real input but need not significantly affect the durability of the structure. Lighter walls may be cheaper than heavy ones but not necessarily less durable. And does it really matter that an eight-inch hand-hewn beam will last for 1000 years if a two-inch beam will last for 200?

II. Implications for the Future

It is clear that some future revision of consumer preferences in favor of housing would serve as a powerful lever in raising the level of residential construction. Such an alteration in consumer standards

could, demographic trends aside, arrest or even conceivably reverse past trends in the share of output absorbed by residential capital formation. Among the factors involved in any radical shift in tastes would be a willingness on the part of home owners and tenants to change old for new dwelling units at a more rapid and continuous pace. An accelerated exchange of old for new products would involve a more rapid decline in prices for old dwellings relative to those for new dwellings, with the ultimate prospect of displacing the least desirable part of the housing stock altogether because it would become worthless, or of making it available to groups which otherwise could not afford separate dwelling units.

A change in consumers' attitudes toward housing could conceivably be induced or strengthened by product innovation. Larger numbers of households might trade old for new units if the latter were vastly superior in design, style, location, or quality, and if the process of innovation could be kept alive so as to be sufficiently persuasive. In such an event, builders would develop a "replacement market" akin to that for other durable goods, in addition to the market hitherto served.

In the light of historical experience, this would be a revolutionary change in the character of the market for new residential construction. For the study of capital formation in residential real estate suggests that the market for additional residential facilities from 1890 to 1950 has been primarily a growth market with accretions to stock equalling approximately the growth in the number of households. Substantial innovations in style, design, and quality have occurred in the past, and yet they failed to generate a "replacement market."

Plotting the future course of consumer behavior with respect to residential construction is no less hazardous than the projection of aggregate consumer spending. But the probability of an autonomous change in favor of housing on the part of the consumer within the next generation must be questioned. It can hardly be anticipated that there will be no further development of new goods and services which will compete for a place in the family budget. Moreover, the position of housing in the budgets of large numbers of consumers may have suffered from rent control during and after the war. Tenants living in controlled apartments have become accustomed to paying a much smaller proportion of income for rent than was customary before the war and have adjusted their expenditures to this pattern. For these households, a spectacular change in consumers' preferences would be required to induce a demand for new housing that could only be satis-

fied at rents or carrying charges much greater than the controlled rents they have been paying for many years.

As to product innovation, the question is whether recent and prospective technological advances and style changes are of an order different from those experienced during the past sixty years. Is the new house of today and tomorrow, designed as it is for highly mechanized household operations, preferably on one floor, so different in kind from older houses that consumers will be induced to trade old for new units in any quantity and on a continuous basis? Will air-conditioning in new dwelling units become a decisive factor in their favor? Is the housebuilding industry developing production and marketing methods substantially different from those in the past? And at what price and rent levels relative to those of old houses will the new products be injected into the market?[9]

In the absence of public action the one factor now visible that portends a possible reversal of past experience is the suburbanization movement. It is conceivable that an intense repugnance towards city residence may lead to absolute declines in central city population. The resulting increase in vacancies and fall in rents and prices might fail to check new construction in the suburbs because the competitive relation between city and suburban housing markets will weaken. Also, suburban living tends to become centered on family and home

9. So far as cost and price reduction of new construction is concerned, a basic difficulty has been pointed out by Ramsey Wood, "Housing Needs and the Housing Market," *Housing, Social Security, and Public Works,* Postwar Economic Series No. 6, Board of Governors of the Federal Reserve System, June 1946, pp. 18–19:

"Proposals to reduce cost to the occupant by reducing particular component costs overlook the important role played by capital values in regulating the amount and kind of new housing added to the supply. It has already been pointed out that new building takes place when the prices of existing houses are higher than the cost of building new houses which consumers regard as comparable, and also that, when building is under way, costs tend to rise to absorb the difference between the market value of existing houses and the sum of the prices of the component resources used in building comparable new houses. This behavior of the market, it has been shown, stops building fairly soon because, when values have risen as high as incomes will permit, and costs have caught up with values, builders cannot operate profitably.

"This process goes on whether the techniques of building and the conditions on which houses are bought and paid for are changing or standing still. The introduction, at strategic times, of techniques and conditions which enable builders or landlords to make new housing available at lower prices than would otherwise be possible may give rise to building which would not have taken place under the earlier circumstances, but the very process of making use of these advantages tends to eliminate them. The market evaluates such advantages against the existing stock of housing and the incomes prevailing in the community; comparable houses command comparable prices; and the cost of building approaches market value."

and may bring about larger and more expensive houses. But the modern suburbanization movement is at least thirty-five years old and, thus far, no such changes have come to pass.

III. The Role of Government

Finally, the profound and increasing interest of government in housing must be taken into consideration. In a broad sense the concern of government with housing and its financing, while it originated in depression emergencies and was intensified by war and postwar dislocations, must be viewed as a lasting and probably increasing influence on residential construction. This concern, by no means limited to housing conditions of the poor and indigent, expresses deep-seated social forces which cause housing to be clothed with substantial and probably growing public interest, although the scope and means of government action may differ in varying political and economic climates. Continued and probably intensified use of federal aids for residential construction is likely not only because they are tools in a broad program to improve housing conditions but also because they fit in with full employment policies.

Thus, the level of residential construction during the next few decades will depend on political as well as on the market-oriented decisions which were controlling before the thirties. Government activity will attempt to maintain a high volume of housebuilding taking the form not only of new dwelling units but of qualitative improvement or renewal of the existing inventory even in the face of declining market demand. To an important degree, the community-at-large through the federal government as its agent will attempt to revise the allocation of real resources to housing — a revision which consumers individually seem to have been unwilling or unable to undertake. Even without a "grand design" in allocating national resources, the rapidly growing interest of the federal government in the support of residential construction may broadly be interpreted as an effort to counteract the results of the historical forces that have led to a relative displacement of housing in the nation's total product.

Past experience is not conclusive as to the effectiveness of such an effort, particularly under adverse conditions such as contractions in employment and income. Consumers' reactions to further liberalization of credit is uncertain, although it is fairly well established that housing demand responds much more to income than to price changes (including changes in the cost of borrowing). This much of a conclusion seems warranted, however: Government policies that result in a larger proportion of total resources being devoted to housing will

probably involve great changes in institutional arrangements for the creation and management of the residential mortgage debt — more drastic perhaps than the changes brought by federal insurance or guarantee of mortgage loans. Liberalization of residential mortgage credit terms under these devices has already gone so far that there are narrow limits to their use for further substantial increases in the ease of borrowing. Thus, demands for "stronger medicine" in the way of federal credit assistance will undoubtedly develop.

The future level of private residential construction will also be affected by government aids to urban redevelopment, which have but recently come into operation on a small scale. In the past, demolitions of residential structures have been so few in relation to the housing stock that their influence on the volume of new housebuilding has been negligible. But government programs to assist in urban redevelopment involve demolition of existing housing and its replacement on a scale far greater than has ever been experienced.

APPENDIX
DERIVATION OF RESIDENTIAL CAPITAL ESTIMATES

The residential capital estimates upon which the per capita and per dwelling units figures depend were obtained by forming a residential wealth estimate for 1890 to which annual estimates of net

TABLE III
COMPARISON OF CUMULATED AND CENSUS-TYPE RESIDENTIAL WEALTH ESTIMATES, VARIOUS DATES 1890–1950
(billions of current dollars)

Date	Cumulated Wealth Estimates[1]			Date	Census-Type Wealth Estimates		
	Structures (1)	Land (2)	Total (3)		Structures (4)	Land (5)	Total (6)
6/1/90	$ 9.0	$ 6.0	$ 15.0	6/1/90[2]	$ 6.7	$ 7.7	$ 14.4
12/31/00	14.6	8.3	22.9	6/1/00[2]	9.5	10.5	20.0
12/31/12	27.3	12.8	40.1	6/1/12[2]	20.7	18.5	39.2
12/31/22	51.1	20.2	71.3	n.a./22[3]	30.0	34.9	65.0
				4/1/30[4]	98.1	24.5	122.6
							(107.6)
12/31/29	80.6	27.9	108.5	n.a./30[3]	51.6	56.1	107.7
12/31/38	75.1	21.7	96.8	n.a./38[3]	44.0	48.0	92.0
12/31/39	77.4	21.8	99.2	4/1/40[5]	—	—	87.4
12/31/49	173.6	38.9	212.4	4/1/50[6]	—	—	260.0

1. Grebler, Blank and Winnick, *op. cit.*
2. Simon Kuznets, *National Product Since 1869*, pp. 201–7. The estimates are derived by successively subtracting from the total taxable real estate estimates made by the various wealth censuses, estimates of agricultural, mining, and manufacturing real estate. The residual, comprising other industrial (commercial) and residential real estate is further subdivided by use of an estimated ratio of residential to the combined residual. The separation of land and structures is performed by dividing each of the categories, total taxable, agricultural, mining, and manufacturing real estate by estimated land ratios. Land and structures for the combined category, other industrial and residential, are then derived as residuals. The effect of this procedure is to assign to

residential real estate the same land ratio as other industrial which is not in accord with available data.

3. Robert R. Doane, *The Anatomy of Wealth*, pp. 213, 224, 251. Doane's procedure is essentially similar to that of Kuznets. Apparently as a result of a typographical error, a 1922 estimate of $67 billion is given on p. 116. Subsequent discussion indicates that $65 billion is the total actually intended. "n. a." indicates month not available.

4. David L. Wickens, *Residential Real Estate*, p. 3. His assumed land ratio is given on p. 4.

5. Bureau of the Census, "Housing-Special Reports," Series H-1943, No. 1, Sept. 11, 1943.

6. Derived as follows from *1950 Census of Housing*, Preliminary Reports, HC-5, No. 1.

15.6 million occupied dwelling units in one-family structures, average value $10,800, total value	= $168.5
17.1 million tenant-occupied dwelling units, estimated average value $3,900, total value (Average monthly rent of $39 multiplied by a factor of 100 to obtain average value.)	= 66.7
3.9 million owner-occupied dwelling units in other than one-family structures, estimated total value (The number of such units represents the difference between all owner-occupied units and owner-occupied single-family houses. The average value was assumed to be equal to the average value of a tenant-occupied unit, namely $3,900.)	= 15.2
2.8 million vacant and other dwelling units, estimated total value	= 9.6
Total	$260.0

capital formation were successively added. The starting estimate of $15.0 billion (current), inclusive of land, represents the product of the total number of nonfarm dwelling units (taken from Wickens' *Residential Real Estate*) and the average value per dwelling unit. The average value per dwelling unit was assumed to equal $1,800 or 55 per cent of the average value of an owner-occupied mortgaged home reported by the 1890 Census. In 1940 the known ratio was 63 per cent.

Because both the starting estimate and all independent benchmark estimates include the value of residential sites, some estimate

TABLE IV

DERIVATION OF PER CAPITA AND PER DWELLING UNIT RESIDENTIAL WEALTH
(all dollar figures in billions)

	Resid. Capital (current dollars) (1)	House Cost Index (2)	Deflated Capital (3)	Nonfarm Population (millions) (4)	Nonfarm Dwelling Units (5)
1890	$ 8.6	39.0	$22.1	33.5	8.3
1900	13.7	38.5	35.5	44.8	10.6
1910	24.4	51.4	47.4	59.9	14.3
1920	50.9	92.1	55.3	74.1	17.7
1930	80.6	100.0	80.6	92.6	25.7
1940	77.4	98.0	79.0	101.5	29.7
1950	173.6	204.3	85.0	114.8[1]	35.7[1]

Source: Grebler, Blank and Winnick, *op. cit.*
Note: Values in Table I are derived from unrounded data and are therefore not precisely reproducible from this table.
1. The Census totals were reduced by 10 per cent to adjust for population and dwelling units not covered in the wealth total.

of the latter was required in order to convert a wealth estimate to an estimate of capital and vice versa. Existing annual FHA data indicate the average ratio of land to the total value of a house for the period since 1936. Benchmark ratios were also discovered for 1907 and 1929. An annual land ratio series was derived from these frag-

ments starting at 40 per cent in 1890 and declining gradually to 18 per cent in 1950. The declining trend in land ratios which is in accord with independent observations is due, primarily, to the opening of vast new areas made possible by the new forms of transportation and, secondarily, to the apartment house which is land economizing. The division between land and structures in the independent estimates, given in Table III, does not appear plausible either in level or movement. Comparison of the cumulated and independent wealth figures is, therefore, made only inclusive of land.

The annual estimates of net residential capital formation represent the difference between gross capital formation and capital consumption. The annual series of gross capital formation (the sum of expenditures on new private nonfarm housekeeping dwelling units and the expenditures on additions to and alterations of such units) combine the new estimates — constructed by David M. Blank[1] for the period 1890–1920 and the official Commerce estimates beginning in 1921. A rough allowance for additions and alterations was added to Blank's series.

The annual capital consumption charge represents the sum of an allowance for basic depreciation and obsolescence and an adjustment for the remaining value of demolished units. The basic depreciation charge, over 90 per cent of the capital consumption allowance, was derived by the declining balance method at a rate of 2 per cent per year. Both the form of depreciation and the specific rate adopted were tied to FHA data in which the current value and reproduction costs of a large sample of single-family houses distributed by age were estimated. The sample houses showed a relatively greater value decline in earlier than in later years at an average rate of something under 2 per cent. Certain minor biases in the sample, however, suggested a small elevation in rate. Since this rate was applied to the stock at the beginning of each year, an additional half-year's depreciation was charged to current gross capital formation. Finally, a separate charge for demolished units was obtained as the product of the estimated number of units demolished and the average value of such units.

The cumulated residential wealth estimates are compared with independent estimates of residential wealth in Table III. When Wickens' estimate for 1930 is reduced by $15 billion because of an error in the census reports of that year and a bias in the multiplier he employed to convert average rents of tenant-occupied units into

1. *The Volume of Residential Construction 1889–1950* (New York: National Bureau of Economic Research, 1954), Technical Paper No. 9.

average values, it is seen that there is a large measure of conformity in all years but 1940 and 1950. It is believed that in both 1940 and 1950 a large part of the discrepancy is caused by differences in valuation. The independent estimates for these years are based on owners' estimates of value (and transformation of rent of tenant-occupied units to value figures) while the cumulated series is tied to the movements of a construction cost index. Although our study shows a long-run conformity between the movement of residential construction costs and house market prices, there are many instances of short-run divergences. Furthermore, owners' estimates of value may be less than completely reliable. The cumulated series was preferred to the census-type estimates because the latter vary greatly in origin and derivation. However, one adjustment in the per capita and per dwelling unit values, given in Table I, seems warranted. The cumulated value estimates for 1950 exclude: (1) The value of publicly financed housing, and (2) the value of private housing not arising from capital formation, i.e., the transfer of houses from the farm to the nonfarm category between 1940 and 1950 because of a change in the census definition of a farm or an actual change in use. Both the affected dwelling units and the population residing in them ought to be subtracted from the denominator of each of the two measures. In 1950, it is estimated that about 1,200,000 units out of a stock of 39,000,000 warrant exclusion. On this account both the per dwelling unit and per capita value measures would be raised by 3 per cent (assuming that the average number of persons per dwelling unit in the excluded categories was the same as the over-all average). On the further assumption that official construction estimates were substantially understated during the 1940–1950 decade, the cumulative estimate for 1950 might be raised by an additional 5 or 6 per cent. Combining these two adjustments, the 1950 values might well be 10 per cent higher than those given in Table I.

Louis Winnick.

Institute for Land Use and Housing Studies
Columbia University

A BEHAVIORAL MODEL OF RATIONAL CHOICE

By Herbert A. Simon*

Traditional economic theory postulates an "economic man," who, in the course of being "economic" is also "rational." This man is assumed to have knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous. He is assumed also to have a well-organized and stable system of preferences, and a skill in computation that enables him to calculate, for the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale.

Recent developments in economics, and particularly in the theory of the business firm, have raised great doubts as to whether this schematized model of economic man provides a suitable foundation on which to erect a theory — whether it be a theory of how firms *do* behave, or of how they "should" rationally behave. It is not the purpose of this paper to discuss these doubts, or to determine whether they are justified. Rather, I shall assume that the concept of "economic man" (and, I might add, of his brother "administrative man") is in need of fairly drastic revision, and shall put forth some suggestions as to the direction the revision might take.

Broadly stated, the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist. One is tempted to turn

to the literature of psychology for the answer. Psychologists have certainly been concerned with rational behavior, particularly in their interest in learning phenomena. But the distance is so great between our present psychological knowledge of the learning and choice processes and the kinds of knowledge needed for economic and administrative theory that a marking stone placed halfway between might help travellers from both directions to keep to their courses.

Lacking the kinds of empirical knowledge of the decisional processes that will be required for a definitive theory, the hard facts of the actual world can, at the present stage, enter the theory only in a relatively unsystematic and unrigorous way. But none of us is completely innocent of acquaintance with the gross characteristics of human choice, or of the broad features of the environment in which this choice takes place. I shall feel free to call on this common experience as a source of the hypotheses needed for the theory about the nature of man and his world.

The problem can be approached initially either by inquiring into the properties of the choosing organism, or by inquiring into the environment of choice. In this paper, I shall take the former approach. I propose, in a sequel, to deal with the characteristics of the environment and the interrelations of environment and organism.

The present paper, then, attempts to include explicitly some of the properties of the choosing organism as elements in defining what is meant by rational behavior in specific situations and in selecting a rational behavior in terms of such a definition. In part, this involves making more explicit what is already implicit in some of the recent work on the problem — that the state of information may as well be regarded as a characteristic of the decision-maker as a characteristic of his environment. In part, it involves some new considerations — in particular taking into account the simplifications the choosing organism may deliberately introduce into its model of the situation in order to bring the model within the range of its computing capacity.

I. Some General Features of Rational Choice

The "flavor" of various models of rational choice stems primarily from the specific kinds of assumptions that are introduced as to the "givens" or constraints within which rational adaptation must take place. Among the common constraints — which are not themselves the objects of rational calculation — are (1) the set of alternatives open to choice, (2) the relationships that determine the pay-offs ("satisfactions," "goal attainment") as a function of the alternative that is chosen, and (3) the preference-orderings among pay-offs. The

selection of particular constraints and the rejection of others for incorporation in the model of rational behavior involves implicit assumptions as to what variables the rational organism "controls" — and hence can "optimize" as a means to rational adaptation — and what variables it must take as fixed. It also involves assumptions as to the character of the variables that are fixed. For example, by making different assumptions about the amount of information the organism has with respect to the relations between alternatives and pay-offs, optimization might involve selection of a certain maximum, of an expected value, or a minimax.

Another way of characterizing the givens and the behavior variables is to say that the latter refer to the organism itself, the former to its environment. But if we adopt this viewpoint, we must be prepared to accept the possibility that what we call "the environment" may lie, in part, within the skin of the biological organism. That is, some of the constraints that must be taken as givens in an optimization problem may be physiological and psychological limitations of the organism (biologically defined) itself. For example, the maximum speed at which an organism can move establishes a boundary on the set of its available behavior alternatives. Similarly, limits on computational capacity may be important constraints entering into the definition of rational choice under particular circumstances. We shall explore possible ways of formulating the process of rational choice in situations where we wish to take explicit account of the "internal" as well as the "external" constraints that define the problem of rationality for the organism.

Whether our interests lie in the normative or in the descriptive aspects of rational choice, the construction of models of this kind should prove instructive. Because of the psychological limits of the organism (particularly with respect to computational and predictive ability), actual human rationality-striving can at best be an extremely crude and simplified approximation to the kind of global rationality that is implied, for example, by game-theoretical models. While the approximations that organisms employ may not be the best — even at the levels of computational complexity they are able to handle — it is probable that a great deal can be learned about possible mechanisms from an examination of the schemes of approximation that are actually employed by human and other organisms.

In describing the proposed model, we shall begin with elements it has in common with the more global models, and then proceed to introduce simplifying assumptions and (what is the same thing) approximating procedures.

1.1 *Primitive Terms and Definitions*

Models of rational behavior — both the global kinds usually constructed, and the more limited kinds to be discussed here — generally require some or all of the following elements:

1. A set of *behavior alternatives* (alternatives of choice or decision). In a mathematical model, these can be represented by a point set, A.

2. The subset of *behavior alternatives that the organism "considers"* or "perceives." That is, the organism may make its choice within a set of alternatives more limited than the whole range objectively available to it. The "considered" subset can be represented by a point set $\mathring{A}$, with $\mathring{A}$ included in A ($\mathring{A} \subset A$).

3. *The possible future states of affairs*, or outcomes of choice, represented by a point set, S. (For the moment it is not necessary to distinguish between actual and perceived outcomes.)

4. A "*pay-off*" function, representing the "value" or "utility" placed by the organism upon each of the possible outcomes of choice. The pay-off may be represented by a real function, $V(s)$ defined for all elements, s, of S. For many purposes there is needed only an ordering relation on pairs of elements of S — i.e., a relation that states that s_1 is preferred to s_2 or vice versa — but to avoid unnecessary complications in the present discussion, we will assume that a cardinal utility, $V(s)$, has been defined.

5. *Information as to which outcomes in S will actually occur* if a particular alternative, a, in A (or in $\mathring{A}$) is chosen. This information may be incomplete — that is, there may be more than one possible outcome, s, for each behavior alternative, a. We represent the information, then, by a mapping of each element, a, in A upon a subset, S_a — the set of outcomes that may ensue if a is the chosen behavior alternative.

6. *Information as to the probability that a particular outcome will ensue* if a particular behavior alternative is chosen. This is a more precise kind of information than that postulated in (5), for it associates with each element, s, in the set S_a, a probability, $P_a(s)$ — the probability that s will occur if a is chosen. The probability $P_a(s)$ is a real, non-negative function with $\sum_{S_a} P_a(s) = 1$.

Attention is directed to the threefold distinction drawn by the definitions among the set of behavior alternatives, A, the set of outcomes or future states of affairs, S, and the pay-off, V. In the ordinary representation of a game, in reduced form, by its pay-off matrix, the set S corresponds to the cells of the matrix, the set A to the

strategies of the first player, and the function V to the values in the cells. The set S_a is then the set of cells in the ath row. By keeping in mind this interpretation, the reader may compare the present formulation with "classical" game theory.

1.2 *"Classical" Concepts of Rationality*

With these elements, we can define procedures of rational choice corresponding to the ordinary game-theoretical and probabilistic models.[1]

A. *Max-min Rule.* Assume that whatever alternative is chosen, the worst possible outcome will ensue — the smallest $V(s)$ for s in S_a will be realized. Then select that alternative, a, for which this worst pay-off is as large as possible.

$$\hat{V}(\hat{a}) = \underset{s \epsilon S_{\hat{a}}}{\text{Min}}\ V(s) = \underset{a \epsilon A}{\text{Max}}\ \underset{s \epsilon S_a}{\text{Min}}\ V(s)$$

Instead of the maximum with respect to the set, A, of actual alternatives, we can substitute the maximum with respect to the set, $\overset{\circ}{A}$, of "considered" alternatives. The probability distribution of outcomes, (6) does not play any role in the max-min rule.

B. *Probabilistic Rule.* Maximize the expected value of $V(s)$ for the (assumed known) probability distribution, $P_a(s)$.

$$\hat{V}(\hat{a}) = \underset{s \epsilon S_{\hat{a}}}{\Sigma}\ V(s)P_{\hat{a}}(s) = \underset{a \epsilon A}{\text{Max}}\ \underset{s \epsilon S_a}{\Sigma}\ V(s)P_a(s)$$

C. *Certainty Rule.* Given the information that each a in A (or in $\overset{\circ}{A}$) maps upon a specified s_a in S, select the behavior alternative whose outcome has the largest pay-off.

$$\hat{V}(\hat{a}) = V(S_{\hat{a}}) = \underset{a \epsilon A}{\text{Max}}\ V(S_a)$$

II. THE ESSENTIAL SIMPLIFICATIONS

If we examine closely the "classical" concepts of rationality outlined above, we see immediately what severe demands they make upon the choosing organism. The organism must be able to attach definite pay-offs (or at least a definite range of pay-offs) to each possible outcome. This, of course, involves also the ability to specify the exact nature of the outcomes — there is no room in the scheme for "unanticipated consequences." The pay-offs must be completely ordered —

1. See Kenneth J. Arrow, "Alternative Approaches to the Theory of Choice in Risk-Taking Situations," *Econometrica*, XIX, 404–37 (Oct. 1951).

it must always be possible to specify, in a consistent way, that one outcome is better than, as good as, or worse than any other. And, if the certainty or probabilistic rules are employed, either the outcomes of particular alternatives must be known with certainty, or at least it must be possible to attach definite probabilities to outcomes.

My first empirical proposition is that there is a complete lack of evidence that, in actual human choice situations of any complexity, these computations can be, or are in fact, performed. The introspective evidence is certainly clear enough, but we cannot, of course, rule out the possibility that the unconscious is a better decision-maker than the conscious. Nevertheless, in the absence of evidence that the classical concepts do describe the decision-making process, it seems reasonable to examine the possibility that the actual process is quite different from the ones the rules describe.

Our procedure will be to introduce some modifications that appear (on the basis of casual empiricism) to correspond to observed behavior processes in humans, and that lead to substantial computational simplifications in the making of a choice. There is no implication that human beings use all of these modifications and simplifications all the time. Nor is this the place to attempt the formidable empirical task of determining the extent to which, and the circumstances under which humans actually employ these simplifications. The point is rather that these are procedures which appear often to be employed by human beings in complex choice situations to find an approximate model of manageable proportions.

2.1 *"Simple" Pay-off Functions*

One route to simplification is to assume that $V(s)$ necessarily assumes one of two values, $(1, 0)$, or of three values, $(1, 0, -1)$, for all s in S. Depending on the circumstances, we might want to interpret these values, as (a) (satisfactory or unsatisfactory), or (b) (win, draw or lose).

As an example of (b), let S represent the possible positions in a chess game at White's 20th move. Then a $(+1)$ position is one in which White possesses a strategy leading to a win whatever Black does. A (0) position is one in which White can enforce a draw, but not a win. A (-1) position is one in which Black can force a win.

As an example of (a) let S represent possible prices for a house an individual is selling. He may regard $15,000 as an "acceptable" price, anything over this amount as "satisfactory," anything less as "unsatisfactory." In psychological theory we would fix the boundary at the "aspiration level"; in economic theory we would fix the bound-

ary at the price which evokes indifference between selling and not selling (an opportunity cost concept).

The objection may be raised that, although $16,000 and $25,000 are both "very satisfactory" prices for the house, a rational individual would prefer to sell at the higher price, and hence, that the simple pay-off function is an inadequate representation of the choice situation. The objection may be answered in several different ways, each answer corresponding to a class of situations in which the simple function might be appropriate.

First, the individual may not be confronted simultaneously with
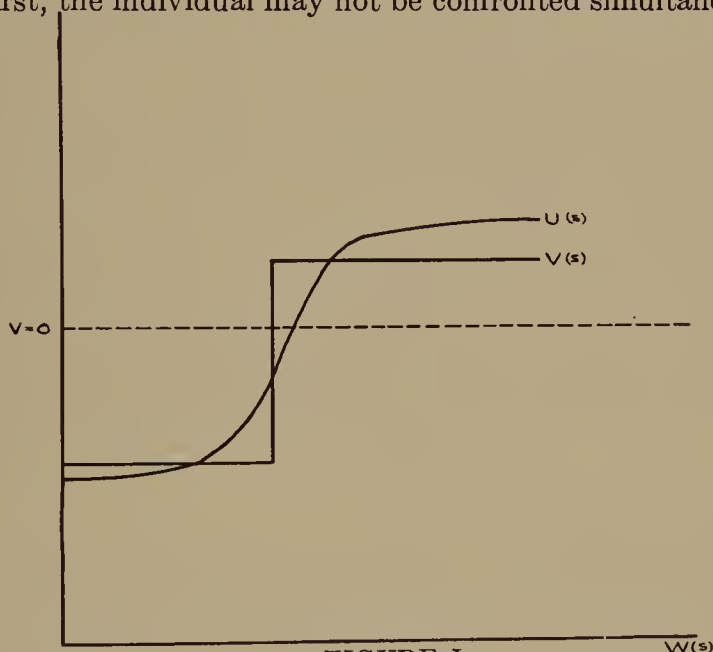
FIGURE I

a number of buyers offering to purchase the house at different prices, but may receive a sequence of offers, and may have to decide to accept or reject each one before he receives the next. (Or, more generally, he may receive a sequence of pairs or triplets or n-tuples of offers, and may have to decide whether to accept the highest of an n-tuple before the next n-tuple is received.) Then, if the elements S correspond to n-tuples of offers, $V(s)$ would be 1 whenever the highest offer in the n-tuple exceeded the "acceptance price" the seller had determined upon at that time. We can then raise the further question of what would be a rational process for determining the acceptance price.[2]

2. See the Appendix. It might be remarked here that the simple risk function, introduced by Wald to bring problems in statistical decision theory within the bounds of computability, is an example of a simple pay-off function as that term is defined here.

Second, even if there were a more general pay-off function, $W(s)$, capable of assuming more than two different values, the simplified $V(s)$ might be a satisfactory approximation to $W(s)$. Suppose, for example, that there were some way of introducing a cardinal utility function, defined over S, say $U(s)$. Suppose further that $U(W)$ is a monotonic increasing function with a strongly negative second derivative (decreasing marginal utility). Then $V(s) = V\{W(s)\}$ might be the approximation as shown on page 107.

When a simple $V(s)$, assuming only the values $(+1, 0)$ is admissible, under the circumstances just discussed or under other circumstances, then a (fourth) rational decision-process could be defined as follows:

D. (i) Search for a set of possible outcomes (a subset, S' in S) such that the pay-off is satisfactory ($V(s) = 1$) for all these possible outcomes (for all s in S').

(ii) Search for a behavior alternative (an a in $\overset{\circ}{A}$) whose possible outcomes all are in S' (such that a maps upon a set, S_a, that is contained in S').

If a behavior alternative can be found by this procedure, then a satisfactory outcome is assured. The procedure does not, of course, guarantee the existence or uniqueness of an a with the desired properties.

2.2 *Information Gathering*

One element of realism we may wish to introduce is that, while $V(s)$ may be known in advance, the mapping of A on subsets of S may not. In the extreme case, at the outset each element, a, may be mapped on the whole set, S. We may then introduce into the decision-making process information-gathering steps that produce a more precise mapping of the various elements of A on nonidentical subsets of S. If the information-gathering process is not costless, then one element in the decision will be the determination of how far the mapping is to be refined.

Now in the case of the simple pay-off functions, $(+1, 0)$, the information-gathering process can be streamlined in an important respect. First, we suppose that the individual has initially a very coarse mapping of A on S. Second, he looks for an S' in S such that $V(s) = 1$ for s in S'. Third, he gathers information to refine that part of the mapping of A on S in which elements of S' are involved. Fourth, having refined the mapping, he looks for an a that maps on to a subset of S'.

Under favorable circumstances, this procedure may require the

individual to gather only a small amount of information — an insignificant part of the whole mapping of elements of A on individual elements of S. If the search for an a having the desirable properties is successful, he is certain that he cannot better his choice by securing additional information.[3]

It appears that the decision process just described is one of the important means employed by chess players to select a move in the middle and end game. Let A be the set of moves available to White on his 20th move. Let S be a set of positions that might be reached, say, by the 30th move. Let S' be some subset of S that consists of clearly "won" positions. From a very rough knowledge of the mapping of A on S, White tentatively selects a move, a, that (if Black plays in a certain way) maps on S'. By then considering alternative replies for Black, White "explores" the whole mapping of a. His exploration may lead to points, s, that are not in S', but which are now recognized also as winning positions. These can be adjoined to S'. On the other hand, a sequence may be discovered that permits Black to bring about a position that is clearly not "won" for White. Then White may reject the original point, a, and try another.

Whether this procedure leads to any essential simplification of the computation depends on certain empirical facts about the game. Clearly all positions can be categorized as "won," "lost," or "drawn" in an objective sense. But from the standpoint of the player, positions may be categorized as "clearly won," "clearly lost," "clearly drawn," "won or drawn," "drawn or lost," and so forth — depending on the adequacy of this mapping. If the "clearly won" positions represent a significant subset of the objectively "won" positions, then the combinatorics involved in seeing whether a position can be transformed into a clearly won position, for all possible replies by Black, may not be unmanageable.[4] The advantage of this procedure over the more common notion (which may, however, be applicable in the opening) of a general valuation function for positions, taking on values from -1 to 1, is that it implies much less complex and subtle evaluation criteria. All that is required is that the evaluation func-

3. This procedure also dispenses with the necessity of estimating explicitly the cost of obtaining additional information. For further discussion of this point see the comments on dynamics in the last section of this paper.

4. I have estimated roughly the actual degree of simplification that might be realized in the middle game in chess by experimentation with two middle-game positions. A sequence of sixteen moves, eight by each player, might be expected to yield a total of about 10^{24} (one septilion) legally permissible variations. By following the general kind of program just described, it was possible to reduce the number of lines of play examined in each of these positions to fewer than 100 variations — a rather spectacular simplification of the choice problem.

tion be reasonably sensitive in detecting when a position in one of the three states — won, lost, or drawn — has been transformed into a position in another state. The player, instead of seeking for a "best" move, needs only to look for a "good" move.

We see that, by the introduction of a simple pay-off function and of a process for gradually improving the mapping of behavior alternatives upon possible outcomes, the process of reaching a rational decision may be drastically simplified from a computational standpoint. In the theory and practice of linear programming, the distinction is commonly drawn between computations to determine the *feasibility* of a program, and computations to discover the *optimal* program. Feasibility testing consists in determining whether a program satisfies certain linear inequalities that are given at the outset. For example, a mobilization plan may take as given the maximum work force and the steel-making capacity of the economy. Then a feasible program is one that does not require a work force or steel-making facilities exceeding the given limits.

An optimal program is that one of the feasible programs which maximizes a given pay-off function. If, instead of requiring that the pay-off be maximized, we require only that the pay-off exceed some given amount, then we can find a program that satisfies this requirement by the usual methods of feasibility testing. The pay-off requirement is represented simply by an additional linear inequality that must be satisfied. Once this requirement is met, it is not necessary to determine whether there exists an alternative plan with a still higher pay-off.

For all practical purposes, this procedure may represent a sufficient approach to optimization, provided the minimum required pay-off can be set "reasonably." In later sections of this paper we will discuss how this might be done, and we shall show also how the scheme can be extended to vector pay-off functions with multiple components (Optimization requires, of course, a complete ordering of pay-offs).

2.3 *Partial Ordering of Pay-Offs*

The classical theory does not tolerate the incomparability of oranges and apples. It requires a scalar pay-off function, that is, a complete ordering of pay-offs. Instead of a scalar pay-off function, $V(s)$, we might have a vector function, V(s); where V has the components $V_1, V_2, \ldots$ A vector pay-off function may be introduced to handle a number of situations:

1. In the case of a decision to be made by a *group of persons*, components may represent the pay-off functions of the individual

members of the group. What is preferred by one may not be preferred by the others.

2. In the case of an individual, he may be trying to implement a number of *values that do not have a common denominator* — e.g., he compares two jobs in terms of salary, climate, pleasantness of work, prestige, etc.;

3. Where each behavior alternative, a, maps on a set of n *possible consequences*, S_a, we may replace the model by one in which each alternative maps on a single consequence, but each consequence has
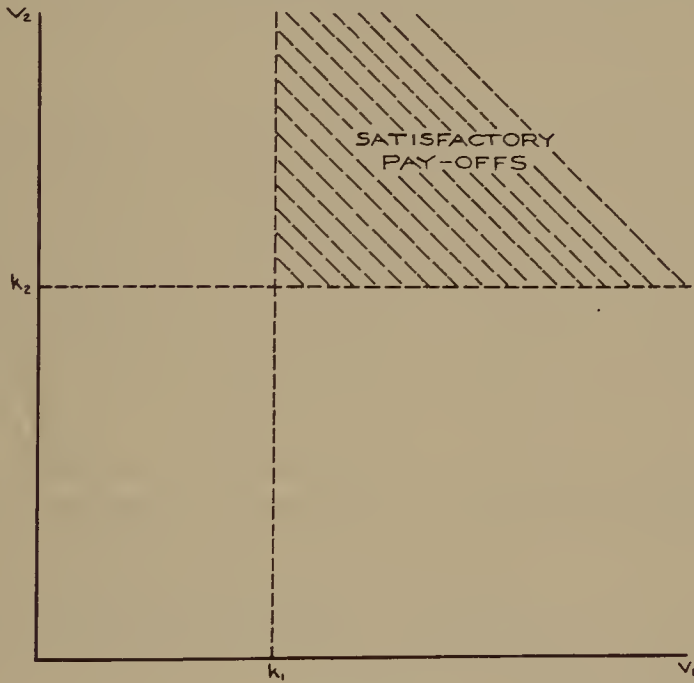
FIGURE II

PARTIAL ORDERING OF PAY-OFFS

as its pay-off the n-dimensional vector whose components are the pay-offs of the elements of S_a.

This representation exhibits a striking similarity among these three important cases where the traditional maximizing model breaks down for lack of a complete ordering of the pay-offs. The first case has never been satisfactorily treated — the theory of the n-person game is the most ambitious attempt to deal with it, and the so-called "weak welfare principles" of economic theory are attempts to avoid it. The second case is usually handled by superimposing a complete ordering on the points in the vector space ("indifference curves"). The third case has been handled by introducing probabilities as weights

for summing the vector components, or by using principles like minimaxing satisfaction or regret.

An extension of the notion of a simplified pay-off function permits us to treat all three cases in much the same fashion. Suppose we regard a pay-off as *satisfactory* provided that $V_i \geqslant k_i$ for all i. Then a reasonable decision rule is the following:

E. Search for a subset S' in S such that $V(s)$ is satisfactory for all s in S' (i.e., $\underset{s \epsilon S'}{V(s)} \geqslant k$).

Then search for an a in A such that S_a lies in S'.

Again existence and uniqueness of solutions are not guaranteed. Rule E is illustrated in Figure II for the case of a 2-component pay-off vector.

In the first of the three cases mentioned above, the satisfactory pay-off corresponds to what I have called a *viable* solution in "A Formal Theory of the Employment Relation" and "A Comparison of Organization Theories."[5] In the second case, the components of V define the *aspiration levels* with respect to several components of pay-off. In the third case (in this case it is most plausible to assume that all the components of k are equal), k_i may be interpreted as the *minimum guaranteed pay-off* — also an aspiration level concept.

III. Existence and Uniqueness of Solutions

Throughout our discussion we have admitted decision procedures that do not guarantee the existence or uniqueness of solutions. This was done in order to construct a model that parallels as nearly as possible the decision procedures that appear to be used by humans in complex decision-making settings. We now proceed to add supplementary rules to fill this gap.

3.1 *Obtaining a Unique Solution*

In most global models of rational choice, all alternatives are evaluated before a choice is made. In actual human decision-making, alternatives are often examined sequentially. We may, or may not, know the mechanism that determines the order of procedure. When alternatives are examined sequentially, we may regard the first satisfactory alternative that is evaluated as such as the one actually selected.

If a chess player finds an alternative that leads to a forced mate for his opponent, he generally adopts this alternative without worry-

5. *Econometrica*, XIX (July 1951), 293–305 and *Review of Economic Studies*, XX (1952–53, No. 1), 40–49.

ing about whether another alternative also leads to a forced mate. In this case we would find it very hard to predict which alternative would be chosen, for we have no theory that predicts the order in which alternatives will be examined. But in another case discussed above — the sale of a house — the environment presents the seller with alternatives in a definite sequence, and the selection of the *first* satisfactory alternative has precise meaning.

However, there are certain dynamic considerations, having a good psychological foundation, that we should introduce at this point. Let us consider, instead of a single static choice situation, a sequence of such situations. The *aspiration level*, which defines a satisfactory alternative, may change from point to point in this sequence of trials. A vague principle would be that as the individual, in his exploration of alternatives, finds it *easy* to discover satisfactory alternatives, his aspiration level rises; as he finds it *difficult* to discover satisfactory alternatives, his aspiration level falls. Perhaps it would be possible to express the ease or difficulty of exploration in terms of the cost of obtaining better information about the mapping of A on S, or the combinatorial magnitude of the task of refining this mapping. There are a number of ways in which this process could be defined formally.

Such changes in aspiration level would tend to bring about a "near-uniqueness" of the satisfactory solutions and would also tend to guarantee the existence of satisfactory solutions. For the failure to discover a solution would depress the aspiration level and bring satisfactory solutions into existence.

3.2 *Existence of Solutions: Further Possibilities*

We have already discussed one mechanism by which the existence of solutions, in the long run, is assured. There is another way of representing the processes already described. Up to this point little use has been made of the distinction between A, the set of behavior alternatives, and $\overset{\circ}{A}$, the set of behavior alternatives that the organism considers. Suppose now that the latter is a proper subset of the former. Then, the failure to find a satisfactory alternative in $\overset{\circ}{A}$ may lead to a search for additional alternatives in A that can be adjoined to $\overset{\circ}{A}$.[6] This procedure is simply an elaboration of the information-

6. I might mention that, in the spirit of crude empiricism, I have presented a number of students and friends with a problem involving a multiple pay-off — in which the pay-off depends violently upon a very contingent and uncertain event — and have found them extremely reluctant to restrict themselves to a set of behavior alternatives allowed by the problem. They were averse to an alternative that promised very large profit or ruin, where the relevant probability could not be computed, and tried to invent new alternatives whose pay-offs were less sensitive to the contingent event. The problem in question is Modigliani's "hot-

gathering process previously described. (We can regard the elements of A that are not in $\overset{\circ}{A}$ as elements that are initially mapped on the whole set, S.)

In one organism, dynamic adjustment over a sequence of choices may depend primarily upon adjustments of the aspiration level. In another organism, the adjustments may be primarily in the set $\overset{\circ}{A}$: if satisfactory alternatives are discovered easily, $\overset{\circ}{A}$ narrows; if it becomes difficult to find satisfactory alternatives, $\overset{\circ}{A}$ broadens. The more *persistent* the organism, the greater the role played by the adjustment of $\overset{\circ}{A}$, relative to the role played by the adjustment of the aspiration level. (It is possible, of course, and even probable, that there is an asymmetry between adjustments upward and downward.)

If the pay-off were measurable in money or utility terms, and *if* the cost of discovering alternatives were similarly measurable, we could replace the partial ordering of alternatives exhibited in Figure II by a complete ordering (an ordering in terms of a weighted sum of the pay-off and the cost of discovering alternatives). Then we could speak of the optimal degree of persistence in behavior — we could say that the more persistent organism was more rational than the other, or vice versa. But the central argument of the present paper is that the behaving organism does *not* in general know these costs, nor does it have a set of weights for comparing the components of a multiple pay-off. It is precisely because of these limitations on its knowledge and capabilities that the less global models of rationality described here are significant and useful. The question of how it is to behave "rationally," given these limitations, is distinct from the question of how its capabilities could be increased to permit action that would be more "rational" judged from the mountain-top of a more complete model.[7]

The two viewpoints are not, of course, completely different, much less antithetical. We have already pointed out that the organism may possess a whole hierarchy of rational mechanisms — that, for example, the aspiration level itself may be subject to an adjustment process that is rational in some dynamic sense. Moreover, in many situations we may be interested in the precise question of whether one decision-making procedure is more rational than another, and to answer this question we will usually have to construct a broader criterion of rationality that encompasses both procedures as approximations. Our whole point is that it is important to make explicit what level we are considering in such a hierarchy of models, and that

dog stand" problem described in *American Economic Review, Proceedings*, XXXIX (1949), 201–8.

7. One might add: "or judged in terms of the survival value of its choice mechanism."

for many purposes we are interested in models of "limited" rationality rather than models of relatively "global" rationality.

IV. Further Comments on Dynamics

The models thus far discussed are dynamic only in a very special sense: the aspiration level at time t depends upon the previous history of the system (previous aspiration levels and previous levels of attainment). Another kind of dynamic linkage might be very important. The pay-offs in a particular trial might depend not only on the alternative chosen in that trial but also on the alternatives chosen in previous trials.

The most direct representation of this situation is to include, as components of a vector pay-off function, the pay-offs for the whole sequence of trials. But then optimization would require the selection, at the beginning of the sequence, of a strategy for the whole sequence (see the Appendix). Such a procedure would again rapidly complicate the problem beyond the computational capacity of the organism. A possible middle ground is to define for each trial a pay-off function with two components. One would be the "immediate" pay-off (consumption), the other, the "position" in which the organism is left for future trials (saving, liquidity).

Let us consider a chess game in which the players are paid off at the end of each ten moves in proportion to arbitrarily assigned values of their pieces left on the board (say, queen, 1; rook, 10; etc.). Then a player could adopt some kind of planning horizon and include in his estimated pay-off the "goodness" of his position at the planning horizon. A comparable notion in economics is that of the depreciated value of an asset at the planning horizon. To compute such a value precisely would require the player actually to carry his strategy beyond the horizon. If there is time-discounting of pay-offs, this has the advantage of reducing the importance of errors in estimating these depreciated values. (Time-discounting may sometimes be essential in order to assure convergence of the summed pay-offs.)

It is easy to conjure up other dynamic complications, which may be of considerable practical importance. Two more may be mentioned — without attempting to incorporate them formally. The consequences that the organism experiences may change the pay-off function — it doesn't know how well it likes cheese until it has eaten cheese. Likewise, one method for refining the mapping of A on S may be to select a particular alternative and experience its consequences. In these cases, one of the elements of the pay-off associated with a particular alternative is the information that is gathered about the mapping or about the pay-off function.

V. Conclusion

The aim of this paper has been to construct definitions of "rational choice" that are modeled more closely upon the actual decision processes in the behavior of organisms than definitions heretofore proposed. We have outlined a fairly complete model for the static case, and have described one extension of this model into dynamics. As has been indicated in the last section, a great deal remains to be done before we can handle realistically a more completely dynamic system.

In the introduction, it was suggested that definitions of this kind might have normative as well as descriptive value. In particular, they may suggest approaches to rational choice in areas that appear to be far beyond the capacities of existing or prospective computing equipment. The comparison of the I.Q. of a computer with that of a human being is very difficult. If one were to factor the scores made by each on a comprehensive intelligence test, one would undoubtedly find that in those factors on which the one scored as a genius the other would appear a moron — and conversely. A survey of possible definitions of rationality might suggest directions for the design and use of computing equipment with reasonably good scores on some of the factors of intelligence in which present computers are moronic.

The broader aim, however, in constructing these definitions of "approximate" rationality is to provide some materials for the construction of a theory of the behavior of a human individual or of groups of individuals who are making decisions in an organizational context. The apparent paradox to be faced is that the economic theory of the firm and the theory of administration attempt to deal with human behavior in situations in which that behavior is at least "intendedly" rational; while, at the same time, it can be shown that if we assume the global kinds of rationality of the classical theory the problems of internal structure of the firm or other organization largely disappear.[8] The paradox vanishes, and the outlines of theory begin to emerge when we substitute for "economic man" or "administrative man" a choosing organism of limited knowledge and ability. This organism's simplifications of the real world for purposes of choice introduce discrepancies between the simplified model and the reality; and these discrepancies, in turn, serve to explain many of the phenomena of organizational behavior.

8. See Herbert A. Simon, *Administrative Behavior* (Macmillan, 1947), pp. 39–41, 80–84, 96–102, 240–44.

APPENDIX

EXAMPLE OF RATIONAL DETERMINATION
OF AN ACCEPTABLE PAY-OFF

In the body of this paper, the notion is introduced that rational adjustment may operate at various "levels." That is, the organism may choose rationally within a given set of limits postulated by the model, but it may also undertake to set these limits rationally. The house-selling illustration of Section 2.1 provides an example of this.

We suppose that an individual is selling a house. Each day (or other unit of time) he sets an acceptance price: $d(k)$, say, for the kth day. If he receives one or more offers above this price on the day in question, he accepts the highest offer; if he does not receive an offer above $d(k)$, he retains the house until the next day, and sets a new acceptance price, $d(k + 1)$.

Now, if he has certain information about the probability distribution of offers on each day, he can set the acceptance price so that it will be optimal in the sense that it will maximize the expected value, $V[d(k)]$, of the sales price.

To show this, we proceed as follows. Let $p_k(y)$ be the probability that y will be the highest price offered on the kth day. Then:

$$(A.1) \qquad P_k(d) = \int_{d(k)}^{\infty} p_k(y) dy$$

is the probability that the house will be sold on the kth day if it has not been sold earlier.

$$(A.2) \qquad \mathcal{E}_k(d) = \int_{d(k)}^{\infty} y \, p(y,k) dy$$

will be the expected value received by the seller on the kth day if the house has not been sold earlier. Taking into account the probability that the house will be sold before the kth day,

$$(A.3) \qquad E_k(d) = \mathcal{E}_k(d) \prod_{j=1}^{k-1} (1 - P_j(d))$$

will be the unconditional expected value of the seller's receipts on the kth day; and

$$(A.4) \qquad V\{d(k)\} = \sum_{k=1}^{\infty} E_k(d)$$

will be the expected value of the sales price.

Now we wish to set $d(k)$, for each k, at the level that will maximize (A.4). The k components of the function $d(k)$ are independent. Differentiating V partially with respect to each component, we get:

$$(A.5) \qquad \frac{\partial V}{\partial d(i)} = \sum_{k=1}^{\infty} \frac{\partial E_k(d)}{\partial d(i)} \qquad (i = 1, \ldots, n).$$

But:

$$(A.6) \qquad \frac{\partial E_i(d)}{\partial d(i)} = \frac{\partial \mathcal{E}_i(d)}{\partial d(i)} \prod_{j=1}^{i-1} (1 - P_j(d)), \qquad \text{and}$$

$$(A.7) \qquad \frac{\partial E_k(d)}{\partial d(i)} = \mathcal{E}_k(d) \prod_{\substack{j \neq i \\ j=1}}^{k-1} (1 - P_j(d)) \left(- \frac{\partial P_i(d)}{\partial d(i)} \right) \text{ for } i < k \text{ and}$$

$$(A.8) \qquad \frac{\partial E_k(d)}{\partial d(i)} = 0 \qquad \text{for } i > k.$$

Hence for a maximum:

$$(A.9) \qquad \frac{\partial V}{\partial d(i)} = -d(i)p_i(d) \prod_{j=1}^{i-1} (1 - P_j(d))$$

$$+ \sum_{k=i+1}^{\infty} \mathcal{E}_k(d) \prod_{\substack{j \neq i}}^{k-1} (1 - P_j(d))p_i(d) = 0.$$

Factoring out $p_i(d)$, we obtain, finally:

$$(A.10) \qquad d(i) = \frac{\displaystyle\sum_{k=i+1}^{\infty} \mathcal{E}_k(d) \prod_{j \neq i}^{k-1} (1 - P_j(d))}{\displaystyle\prod_{j=1}^{i-1} (1 - P_j(d))}.$$

$$= \sum_{k=i+1}^{\infty} \mathcal{E}_k(d) \prod_{j=i+1}^{k-1} (1 - P_j(d)).$$

For the answer to be meaningful, the infinite sum in (A.10) must converge. If we look at the definition (A.2) for $\mathcal{E}_k(d)$ we see this

would come about if the probability distribution of offers shifts downward through time with sufficient rapidity. Such a shift might correspond to (a) expectations of falling prices, or (b) interpretation of y as the *present value* of the future price, discounted at a sufficiently high interest rate.

Alternatively, we can avoid the question of convergence by assuming a reservation price $a(n)$, for the nth day, which is low enough so that $P_n(d)$ is unity. We shall take this last alternative, but before proceeding, we wish to interpret the equation (A.10). Equation (A.10) says that the rational acceptance price on the ith day, $d(i)$, is equal to the expected value of the sales price if the house is not sold on the ith day and acceptance prices are set optimally for subsequent days. This can be seen by observing that the right-hand side of (A.10) is the same as the right-hand side of (A.4) but with the summation extending from $k = (i + 1)$ instead of from $(k = 1)$.[9]

Hence, in the case where the summation is terminated at period n — that is, the house will be sold with certainty in period n if it has not been sold previously — we can compute the optimal $d(i)$ by working backward from the terminal period, and without the necessity of solving simultaneously the equations (A.10).

It is interesting to observe what additional information the seller needs in order to determine the rational acceptance price, over and above the information he needs once the acceptance price is set. He needs, in fact, virtually complete information as to the probability distribution of offers for all relevant subsequent time periods.

Now the seller who does not have this information, and who will be satisfied with a more bumbling kind of rationality, will make approximations to avoid using the information he doesn't have. First, he will probably limit the planning horizon by assuming a price at which he can certainly sell and will be willing to sell in the nth time period. Second, he will set his initial acceptance price quite high, watch the distribution of offers he receives, and gradually and approximately adjust his acceptance price downward or upward until

9. Equation (A.10) appears to have been arrived at independently by D. A. Darling and W. M. Kincaid. See their abstract, "An Inventory Problem," in the *Journal of Operations Research Society of America*, I, 80 (Feb. 1953).

he receives an offer he accepts — without ever making probability calculations.   This, I submit, is the kind of rational adjustment that humans find "good enough" and are capable of exercising in a wide range of practical circumstances.

<div align="right">Herbert A. Simon.</div>

Carnegie Institute of Technology

THE KINKED OLIGOPOLY CURVE RECONSIDERED*

By C. W. EFROYMSON

In his article, "The Kinky Oligopoly Demand Curve and Rigid Prices,"[1] Professor George J. Stigler, after examining the "pure theory" of the kinked curve and "the degree of correspondence between the price patterns implied by the theory and the observed price patterns in oligopolistic industries," concludes that the "empirical evidence reveals neither price experiences that would lead oligopolists to believe in the existence of a kink nor the pattern of changes of price quotations that the theory leads us to expect."[2] This article has been widely noticed but has not been subjected (to my knowledge) to detailed criticism. I shall criticize Stigler's arguments and evidence, maintaining that both are deficient, and that his "adverse conclusion" is not justified.

I. BELIEF IN THE EXISTENCE OF THE KINK

In his analysis of "The Workings of the Kink,"[3] Stigler describes the (hypothetical) experience of duopolist A, who, after demand for his product has increased sharply, finds it profitable to raise price even though he assumes that rival B will not follow. If, however, "B's costs and demand are similar to A's," B, he argues, will raise his price "simultaneously."[4] By such experience, Stigler thinks that

---

1. *The Journal of Political Economy*, LV (1947), 432–49, reprinted in *Readings in Price Theory*, eds. G. J. Stigler and K. E. Boulding (published for the American Economic Association; Chicago: Richard D. Irwin, Inc., 1952), pp. 410–39.

2. *Ibid.*, pp. 410, 434–35. (Page references to Stigler's article refer to the reprinted article.)

3. *Ibid.*, pp. 416–18.

4. *Ibid.*, p. 417.

"the existence of the kink is contradicted," and "businessmen will learn . . . that there is no kink."[5]

I propose now to re-examine Stigler's example. It is important to observe that, according to Stigler, after the increase of demand from D_1 to D_2, which expands sales and moves the profit maximizing price of each seller away from the kink,[6] A continued to conjecture the same reactions of his rival B which determined the original demand curve. The new curve D_2 is still (obtusely)[7] kinked, and at the orig-

inal equilibrium price p_o. This conjecture (which was later proven erroneous by B's action) is, in the given market situation, surprising. After demand had increased so sharply (according to Figure 3, about 50 per cent at the price p_o), A was either operating nearer to least-cost output, or he was beyond it; therefore he was probably not overly hungry for increased volume. Moreover, after the change in demand, he anticipated an increase of profits *even if B did not raise price with him.* If — as a shrewd duopolist — A surmised that B's situation was similar to his own, he ought, it seems, therefore to have guessed that B would *now* be as favorable as he himself was to a price
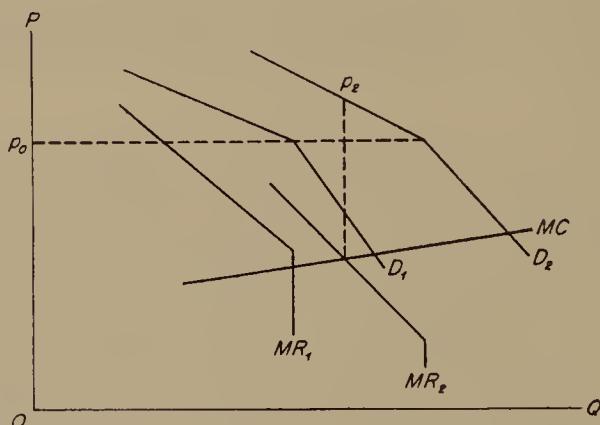
5. *Loc. cit.*

6. *Ibid.*, Figure 3. Reproduced herewith, by courtesy of the Universtiy of Chicago Press. Copyright 1947, the University of Chicago.

7. The kink in the demand curve is *obtuse* if the oligopolist conjectures that rivals will in some measure conform to his downward, but not to his upward price revisions; and *reflex* if he conjectures that rivals will conform to his upward but not to his downward revisions. Cf. my article, "A Note on Kinked Demand Curves," *American Economic Review*, XXXIII (1943), 98–109. Stigler, except for one reference, *op. cit.*, p. 412, n. 4, discusses only obtusely kinked demand functions, which he calls "kinky" without the qualifying adjective. Paul M. Sweezy, "Demand Under Conditions of Oligopoly," *Journal of Political Economy*, XLVII (1939), 568–73, reprinted in *Readings in Price Theory*, describes the reflex curve without so naming it.

increase. If he had so guessed or "conjectured," his demand curve (prior to any "experience") would not have been obtusely kinked. It would not have been, that is, D_2 of Stigler's Figure 3.

But, whatever the reasons, A did not so conjecture, and was surprised, albeit pleasantly, when B raised his price, since his own profits were thus further increased. If, after this experience, A concludes that "there is no kink," he apparently believes that B will henceforth conform both to his (A's) upward and downward price revisions.[8] He will therefore again find it profitable to raise price, and B too will raise his as before. But unlike the situation of the first price move in Stigler's example, a point will soon be reached beyond which either duopolist's profits will fall if he raises his price and the rival does *not* conform. Moreover, B in these new situations is increasingly tempted not to conform by the prospect of short-run advantage if his price is less than A's. The contingent disadvantage to B of not conforming is, of course, the likelihood that A will not long maintain a price higher than his own (B's). But if A continues to raise his price on the basis of the conjecture that B will conform, sooner or later, he will once more be surprised, this time unpleasantly.

It is possible, of course, that, after each experience which surprises him, A will continue with the same conjecture as before — whether conformity or nonconformity — and frequently find himself in error. But it is, I think, more likely that A will reconsider the premises of his conjectures, and conclude that B's reactions to his price revisions depend on the particular market situations; and further perhaps that, both in depressed and exuberant markets, B is likely to follow in one, and not to follow in the other, direction.

In my "Note on Kinked Demand Curves,"[9] I have tried to demonstrate the relationship of conjectures which imply reflexly kinked demand functions to markets of exuberant demand and full use of facilities; and of conjectures which imply obtusely kinked functions to depressed markets and "excess capacity." These relation-

---

8. In this event, A's demand curve is of the nature of Chamberlin's steeper demand function DD' (*The Theory of Monopolistic Competition*, p. 91, Figure 14). But Stigler, *op. cit.*, p. 417, also says: "The situation [when B raises price] becomes classical duopoly — with the usual wide range of possible patterns of behavior." If by "classical duopoly," Stigler means that a situation prevails wherein the duopolist ignores possible reactions of the rival or assumes that the rival will not react (as with Cournot, Bertrand and Edgeworth), his demand curve is of the nature of Chamberlin's function dd' (*loc. cit.*). In this case, if demand declines, A is likely to reduce price, *expecting B not to follow*. But B will surely follow.

9. *Op. cit.*

ships may be summarized as follows: If demand is insistent and firms are operating well beyond their least-cost outputs (rising average total unit costs), oligopolists are not likely to be hungry for increased sales; and perhaps not loath to lose some sales if one of their number cuts prices. An oligopolist may, therefore, conjecture that rivals will follow an upward but not a downward revision of price, which conjectures determine the reflexly kinked demand curve implying volatile prices — the opposite of the implications of obtusely kinked demand curves.[1] To be sure, it is quite possible that, because businessmen are characteristically cautious, and because in their calculations they tend to weigh the fear of losses more heavily than the prospect of extra gains, reflex curves may be of rare occurrence, even in exuberant markets. And even if entrepreneurs make the conjectures which imply reflex curves, prices are likely to be less volatile than indicated because of various other considerations which make for comparative inflexibility of oligopolistic — and many other — prices, irrespective of market conditions.

If, however, demand is slack, the price leader, or any oligopolist, is likely to conjecture that rivals will follow his downward, but not his upward, price revisions. His demand function is then obtusely kinked; and his price is, within limits of certain changes in demand and/or cost functions, stable. The emergence of the obtuse kink, then, with its implied stability of price is neither a necessary result of oligopoly as such, nor a function of the passage of time.[2] It is a result specifically of the presently prevailing market situation, wherein oligopolists, basing conjectures in part at least on their own situations, think that rivals are hungry for increased sales and fearful of reduced sales.[3]

In the second part of his article, "Empirical Tests of the Theory,"[4] Stigler says that "If price reductions are not followed by rivals, or price increases are as closely followed" no "objective basis" for belief in the kink "exists."[5] He relates price histories of seven industries during periods ranging from 1918 to 1937,[6] and shows that, except in

1. *Ibid.*, pp. 104–7.
2. Contrast Stigler, *op. cit.*, pp. 417–18: "the theory does not explain why prices that have once changed should settle down . . . and gradually produce a new kink. One possible explanation might be that a period of stability of demand and costs has created a tradition of stable prices . . ."
3. The foregoing list of possibilities is not exhaustive. Oligopolists may also engage in price warfare, or may "co-ordinate" policies.
4. *Op. cit.*, pp. 420 ff.
5. *Ibid.*, p. 421. Actually the sentence quoted describes a reflex kink.
6. The industries are: Cigarettes (three firms); automobiles (Ford and Plymouth only); anthracite coal (seven firms); steel (United States Steel and

the cigarette industry in 1918, price increases in the seven industries were more closely followed than were decreases. He concludes that: "In these seven industries there is little historical basis for a firm to believe that price increases will not be matched by rivals and that price decreases will be matched . . . here are seven industries in which the existence of the kinky demand curve is questionable."[7]

Stigler here seems to imply that "belief" in the kinked demand curve, and even its "existence" (as a basis of entrepreneurial decision) depend on continuing or repeated corroboration by (unhappy) experiences of the conjectures which it describes. Such ex-post verification may be possible for demand functions[8] in markets wherein prices and quantities supplied change over time even though the supplying firms individually and collectively would prefer no changes, or changes other than those which occur. Thus, by the use of corrected or adjusted time series, experiential verification of estimates of demand is at least theoretically possible.

But with respect to individual firm demand functions of the type developed by Chamberlin,[9] corroboration by experience is unlikely to be available because the quantity supplied changes, *ceteris paribus*,

---

Bethlehem only); dynamite (three firms); gasoline in the Boston area (two firms); potash (number of firms not stated).

7. *Ibid.*, p. 425. The rayon industry, which Stigler cites only in his discussion of price rigidities (see below) affords experience of an opposite nature. Jesse W. Markham, *Competition in the Rayon Industry* (Cambridge: Harvard University Press, 1952), p. 88, says, "On the average rayon producers are slightly more hesitant in following a price increase than a price decrease." But, "since most firms meet their competitors' lower prices immediately regardless of the effective dates they show on their published price lists, the actual difference [greater delay in meeting price increases] is probably about ten days . . ." Stigler apparently uses "announced" prices, or (nominal) prices as reported by the Bureau of Labor Statistics, *op. cit.*, p. 426, for all of his "empirical tests." He discusses deviations of actual (transaction) from nominal prices in relation to price rigidities, *ibid.*, pp. 425–26 (see below, Section II), but, except for steel, does not consider them in relation to his "price histories." Since deviations of actual from nominal prices are always, or predominantly, *downward*, the probability of off-list selling in the seven industries, in addition to steel, implies that actual delays in meeting downward price reductions were in some cases less than the delays measured by Stigler. (Cf. below, p. 131, note 1, deviations of actual from nominal prices in the cigarette industry.)

8. In this section, Stigler does not state whether the demand functions, on the basis of which prices were revised, had shifted, or not shifted. Although the question is probably not very significant in the context, I shall assume — since the problem of belief in or existence of the (earlier) kink after demand has shifted is analyzed above — that demands, during the various periods of price revision, were generally stable.

9. *Op. cit.*

only when the entrepreneur expects thereby to increase profit.   Only one situation[1] described by the demand function is profit-maximizing so long as the supply function is unchanged, and, in contrast to the situation in non-oligopolistic competitive markets, price and supply are determined by the individual entrepreneur.   Kinked demand functions are, in short, ex-ante estimates, expressing conjectures or "beliefs,"[2] which guide conduct by indicating the unfavorable outcomes of all but one of the possible prices.   Moreover, the conjectured inelasticity of demand below, and its elasticity above, the prevailing price are "barriers" to price revisions which might otherwise be profitable; and "business" in Stigler's words is "the collection of devices for circumventing barriers to profits."[3]   Or, at the least, it is not a pursuit wherein oligopolists obstinately raise prices when they believe that rivals will not follow, and reduce them when rivals will follow — unless indeed the underlying conditions have changed so radically that the revision of supply and/or price will be profitable in spite of the conjectured reactions of the rivals.

It is now, I believe, possible to reappraise the price histories cited by Stigler.   With a few significant exceptions, upward price revisions were simultaneous or followed by rivals within a few days.   This indicates that the left segments of the demand curves were — or should have been — based on conjectures that rivals would adjust to, that is, follow, price increases.   The individual demand functions in so far as rivals would have followed downward revisions were apparently[4] unkinked, of the type described by the demand function DD' (not dd' or DQd') in Chamberlin's Figure 14.[5]   But the fact that in these cases, oligopolistic demand curves were proven, as it were by

1. In the exceptional case of the "reflex" demand curve, there are usually two supply quantities both of which indicate greater profits than the profit at the existing supply and price; but even in this event, one of the two quantities would normally indicate a greater profit than the other.   See my article, *op. cit.*, pp. 104–7.

2. Cf. Stigler, *op. cit.*, p. 411, n. 2.   "The demand curve for the product of an oligopolist can be defined only if the reactions of rivals to price changes are known. . . . This demand curve is objective in the sense that it is consistent with the facts . . . if rivals behave in the specified manner."   But, except perhaps in the case of overt collusion with complete assurance that rivals will adhere, (future) reactions of rivals can never be "known."   And, if the conjecturing oligopolist (wisely) does not revise price, rivals will have no occasion to behave in the specified manner.

3. *Ibid.*, p. 435.

4. "Apparently," because it is possible that demand and/or supply functions may, in any case, have shifted in such manner that price revisions were indicated even though rivals might not have followed.

5. *Op. cit.*, p. 91.

the events, to be unkinked does not mean that they are never or only infrequently kinked. It appears, merely, that the oligopolists raised prices when they had reason to believe that rivals would follow. The barrier of the kink, if it had ever existed, had been removed — circumvented.

Just how this circumvention was accomplished in the seven industries is not stated explicitly by Stigler. But it is possible to guess. For instance, the prices of seven anthracite producing companies, Stigler tells us, "for each size of coal . . . are identical and they almost invariably change on the same day."[6] Apparently, in this industry, circumvention was achieved by a well-established habit of price leadership, by quasi or overt agreement. Price leadership and covert or overt collusion are indeed oligopolistic situations which, admittedly in a negative and not wholly satisfactory manner, indicate "belief" in the existence of a kink, or, more exactly, belief that it would prevail if there were among the oligopolistic rivals no understanding or established custom by which it is ironed out.

Experiences in three of the industies cited by Stigler diverge in various respects from the more general pattern. First, American Tobacco Company in September, 1918, raised prices of Lucky Strikes. Rivals did not follow, and the price rise was rescinded in November, following a 31 per cent drop in sales. Subsequently, prices "continued to decline for several months."[7]

This unhappy experience indicates that American Tobacco in September, 1918, expected rivals to follow its increase. The firm, that is, estimated demand on the assumption that Chamberlin's function DQ described demand at prices above the prevailing price BQ and, to its cost, discovered that instead it was dQ.[8] This impressive lesson may have sufficed for some time,[9] and may have convinced

6. *Op. cit.*, pp. 422–23.
7. *Ibid.*, p. 421. William H. Nicholls, *Price Policies in the Cigarette Industry* (Nashville: Vanderbilt University Press, 1951), p. 51, says American's sales in January, 1919, were 53.4 per cent below sales of September, 1918.
8. *Op. cit.*, p. 91.
9. Nicholls, *op. cit.*, p. 176, says that American's experience in 1918 "and its experience with *matched price cuts* during 1920–22 were undoubtedly such as to convince it of the reality of the 'kink' " (italics added); and, p. 57, that: "With the adoption of virtually identical prices in August, 1923, the three firms brought the era of experimentation with price competition to a close. Price competition having been found wanting . . . competition was shifted almost wholly to a non-price basis." Stigler, *op. cit.*, p. 421, lists price changes 1928–1934, and says, "The later price history in this industry, however, is not such as to create a belief in the existence of a kink." I suggest, rather, that the impressive lessons of the earlier period had not been forgotten.

oligopolists of other industries also of the unprofitability of price increases which rivals will refuse to follow.

The second variant experience is in steel. Bethlehem, according to Stigler, "faithfully followed" the price increases of United States Steel. But in this industry "official price lists and trade-journal quotations have been honored almost exclusively in the breach . . ."[1] This refers, of course, to unavowed *downward* price revisions (see above, p. 123, note 7) and the question arises: Why, when they seek increased sales do steel producers — and other oligopolists — sometimes, or in the case of steel in the cited period generally, *prefer* concealed reductions? To this question there are, aside from the necessary delays in printing or distributing revised quotations, two plausible answers: first, the advantage of price discrimination among customers, which has nothing to do with possible belief in the kink; and second, the hope of enjoying larger increases of demand at lower prices which rivals do not recognize or follow. The honoring of price lists "almost exclusively in the breach" thus may be indirect evidence of belief that the right wing of the demand curve is, or would be, if rivals knew or were forced to take cognizance of the reduction, the QD', and not the Qd' of Chamberlin's Figure 14 (cited above). In this case, if the oligopolists think that rivals will not follow upward price revisions, the demand curves are kinked.

The third variant is the case of the American Potash and Chemical Corporation which, on June 26, 1934, withdrew the price reduction of June 1 because, contrary to "long prevailing custom," other potash producers failed to follow. Here is clear evidence of the custom of price leadership — apparently quasi-agreement — by which the kink had been circumvented. But co-ordination of price policy among oligopolistic rivals, whether "spontaneous"[2] or overt, does not permit price leaders full freedom of action. Rebellion is always possible if the price revision of the leader affects followers adversely. In the present case, American Potash apparently expected to increase profit or reduce losses by a general price reduction. Rivals apparently would have suffered by the general reduction, and would not conform. Rather than risk price warfare, American Potash withdrew its reduction. This experience illustrates, I suggest, only the difficulty, and the importance to oligopolists of maintenance of an existing habit of policy co-ordination.

The foregoing, I believe, indicates the following conclusions: (1)

1. *Ibid.*, p. 423.
2. William Fellner, *Competition Among the Few; passim.*

The fact that in the seven industries, price increases or increases of nominal prices (see note 7, p. 123, above) were more frequently and closely followed by rivals than were price decreases indicates that the oligopolists wisely raised prices only when they believed rivals were willing to follow, that is, when they had reason to believe that — by covert or overt agreement — the kink had been ironed out.[3] (2) Experience can corroborate the existence of, or the validity of "belief" in the kinked curve, only when an oligopolist has disbelieved, and has revised price on the basis of his disbelief. Unless oligopolists are more or less obstinately unwise, such experiences are likely to be infrequent; and the lessons of experiences of this sort — as in the American Tobacco case — are likely to be impressive. The exceptional cases, therefore, when price increases have not been closely followed are evidence of *disbelief* of the price-increasing oligopolists in the kink, and also, of course, of the faultiness of their conjectures. Experience is thus available only in exceptional instances to prove the invalidity of the unkinked demand curve. Unfortunately the cases of oligopolistic "belief" in the kink are incapable of analagous experimental — ex post facto — verification, because the oligopolists who believe do not raise prices. (3) Since the kink-anticipated refusal of rivals to follow price increases, or the fear that they will refuse — is a "barrier to profits," the likely result of the kink-barrier is an effort to eliminate or to circumvent it. With respect to downward price revision, secret price concessions, discounts, allowances, etc. are methods of circumvention. Price leadership and collusion are means for the elimination of the kink. And, finally, although it has not yet entered into the discussion, non-price competition and product variation are outflanking movements, as it were, whereby the barrier may be circumvented.

II. Price Patterns: Rigidity

Stigler compares flexibilities of prices between the following industry groups or situations:[4]

1. Two monopolies and nineteen oligopolies.

3. Stigler, *op. cit.*, p. 425, says: "This indicates only that not every oligopoly has reason to believe that it has a kinky demand curve . . ."

4. Stigler also, *ibid.*, Table 2, p. 430, correlates the number of firms in his nineteen oligopolies with price flexibilities. He expects, "on the kink theory," that the kink will disappear with "very few rivals" (2, 3, 4?) and again "with many rivals" (9, 11, 12?); but finds "a weak tendency for a greater number of firms to be associated with a greater frequency and amplitude of price changes, the contrary of the implications of the kinky demand curve." (*Ibid.*, p. 431.) The significance of this finding is, it seems to me, obscure.

2. Five oligopolies during periods of (charged or proven) collusion, and during other periods.

3. Oligopolies with price leadership of the "dominant firm" type,[5] and oligopolies without price leadership.

4. Six oligopolies with heterogeneous products and thirteen oligopolies with homogeneous products.

With respect to each of the four comparisons, or "tests," Stigler states or implies that there is no kink in the demand curves of the first-named group,[6] and that, therefore, prices of these groups ought in each case, according to the theory, to be more flexible than the prices of the second-named group. His measures and indices indicate, however, that prices of the first-named group are in each case less flexible than prices of the second-named group. "The industries included in these tests are not very numerous," he says, "but they are sufficiently varied and important to suggest that similar adverse results would be secured from a larger sample."[7]

These tests, I think, are deficient and inconclusive because: (a) Stigler measures flexibilities of prices of (unco-ordinated) oligopolies for a "complete business cycle," thus including periods when oligopolists did not base pricing decisions on obtusely kinked demand curves; and (b) the data of the comparisons are list or nominal prices, from which transaction prices diverge, at least in co-ordinating oligopolies, in characteristic manners. It should also be pointed out that, (c) with respect to the co-ordinating oligopolies (collusion and price leadership), attainment of agreement does not eliminate the kink if uncertainty that rivals will conform remains; and, finally, (d) with respect to the first and fourth comparisons, as listed above, Stigler's finding of the relative rigidity of prices of monopolies and heterogeneous oligopolies indicates only that (obtuse) kinks are but one of the various factors contributing to price inflexibility.

A. *The Time Period*

Stigler's measures of flexibility are based on the number of changes of price from June, 1929, through May, 1937, "a complete business cycle."[8] Except, therefore, with respect to the five oligop-

---

5. For reasons which I do not understand, Stigler, *ibid.*, p. 433, excludes oligopolies with the "barometric" type of price leadership.

6. With respect to the fourth comparison, heterogeneity of products reduces the significance of rivals' reactions. Kinks may remain, but the degree of price rigidity (length of the discontinuity of the marginal revenue curve) is reduced.

7. *Ibid.*, p. 435.

8. *Ibid.*, p. 426.

olies, which engaged for limited periods in explicit collusion, he makes no distinction between intermediate periods when oligopolists did and did not determine prices on the basis of obtusely kinked demand curves.

Stigler, here, as in his earlier discussion of the re-emergence of the kink, does not relate the (obtuse) kink to the depressed, or non-exuberant, markets in which its occurrence is likely. He thus implies that, according to the theory, oligopolists' prices must *always* or *generally* be less flexible than prices of non-oligopolistic sellers, or of co-ordinating oligopolists. This would be correct if non-co-ordinating oligopolists *always* or *generally* determined prices on the basis of obtusely kinked demand functions, or if, when they did not do so, their prices were not peculiarly unstable.

There are, however, two market situations, both peculiar to oligopolistic competition, and both characterized by exceptional instability of prices: first, oligopolists — and only oligopolists — occasionally engage in price warfare. During these episodes, which are struggles for long-run advantages, short-run demand curves, whether kinked or continuous, are a minor factor in pricing decisions. Second, when markets are exuberant, oligopolists may make those conjectures which determine reflex kinks and, so far as other considerations do not outweigh the kink influence, unstable prices. The inclusion in Stigler's measures[9] of periods when oligopolists were engaged in price warfare or when they conjectured reflex kinks, if such periods occurred, raises the measure of price flexibility of the second-named group in each of the four comparisons.

Stigler does not consider in these tests the possible effects on price flexibility of bilateral complications. In his earlier "Elaboration of the Theory" he discusses briefly the possibility that buyers of the oligopolists' products may be few. "If few buyers face a competitive supply, each will presumably believe that, if he raises his buying price, rivals will follow . . . and, if he lowers his buying price, rivals will not follow . . ."[1] These conjectures are, however, plausible principally when markets are exuberant. In the less insistent markets, when oligopolists conjecture obtusely kinked curves, oligopsonists are more likely to conjecture that rivals will follow price decreases and not follow price increases. Such conjectures determine kinks in supply curves[2] which, like reflex kinks in demand curves, imply

9. *Ibid.*, Tables 1–4, pp. 429–34.
1. *Ibid.*, p. 416, n. 12. Thus, according to Stigler, prices "will be especially rigid . . . The difficulty is that the assumptions are inconsistent."
2. M. Bronfenbrenner, "Some Applications of the Discontinuous Oligopoly

instability of prices. Oligopsonist buyers in this situation are strongly tempted to push prices down, just as and when oligopolists are impelled to hold to the prevailing price. Unless, then, either the buyers or sellers mistake the character of the market, pressures generated by kinks in the demand and in the supply curves of a commodity will be of opposite nature.[3] If the opposing forces are of approximately equal strength, they will, of course, tend to cancel out. But if the oligopsonistic pressure, in buyers' markets, predominates, the prices of the oligopolists, in depressed markets, are likely to be relatively unstable.[4]

Stigler's measures of price flexibility include, then, periods when prices of (unco-ordinated) oligopolies ought, according to the kink theory, to have been relatively stable, and periods when they ought, according to the theory as here outlined, to have been relatively volatile, and also situations wherein prices were, or may have been, determined primarily by monopolists or oligopsonists to whom the oligopolists sold. Bilateral complications and also reflex kinks in the generally depressed markets of 1929–1937 were perhaps of little significance.[5] Episodes of price warfare, however, certainly occurred, and the exceptional volatility of prices during such episodes, however brief, raises the average number of price changes and the indices of price flexibility of some at least of the non-co-ordinating oligopolies of the comparisons. The *possible* comparative rigidity of the prices of the (second-named) oligopolistic groups, which Stigler looked for and did not find, is, therefore, effectively concealed by Stigler's failure to exclude periods when the oligopolists engaged in price warfare or did not make the conjectures which determine obtuse kinks.

B. *Nominal versus Transaction Prices*

Stigler calls attention to his use of "quoted prices on a temporal basis" whereas "we need transaction prices." Referring specifically

Demand Curve," *Journal of Political Economy*, XLVIII (1940), 420–22, briefly discusses kinked supply curves.

3. Similar counter-pressures are engendered when oligopolists are also oligopsonists for significant portions of their supplies. But these relationships are probably of less significance than bilateral complications on the selling side.

4. If markets are exuberant, the two kinks will again be of opposite nature, with oligopolists attempting to change prices and oligopsonists resisting price revision.

5. Opposing bilateral complications were none the less present in some of the industries: e. g., tractor, plow, and can manufacturers purchased steel and tin plate from oligopolists, and buyers of cans may have exerted oligopsonistic pressure.

to deviations of transaction from nominal prices, he says, "We cannot infer that all nominally rigid prices are really flexible, but there is also very little evidence that they are really rigid."[6] This "deficiency" of the data, although it applies generally to all the industry groups of the comparisons, and also to the price histories of Stigler's earlier discussion (see above), is illustrated in glaring fashion by the case of the colluding rayon yarn producers (second comparison). Stigler says "there were no price changes" from October 21, 1931, to May 23, 1932, the period of collusion.[7] Jesse W. Markham, however, points out that "although no changes were made in the list price of rayon between October 21, 1931 and May 23, 1932, 'misbranding' and off-list selling continued and by April . . . even such large producers as Viscose, Dupont, and Industrial were selling rayon at 30 percent below list price."[8] In this case, at least, the rigidity of nominal prices during explicit collusion is trivial.

Despite the "shortcomings" of his data, however, Stigler thinks that the comparisons have "some value. If the theory cannot explain the pattern of rigidity of quoted prices among industries, there is no presumption that it would explain the pattern of transaction price rigidity among industries."[9] Stigler here overlooks the specific relationship of off-list selling to oligopoly, and particularly to oligopolies when policies are co-ordinated. Monopolists and also oligopolists may, as stated, sell below list to gain the extra profits of discrimination, and also because of the expenses and necessary delays in issuing new lists. But only oligopolists resort to unavowed price concessions in order to avoid informing *rivals*, and only co-ordinating oligopolists deceive "price collectors" in order to "strengthen morale within the industry."[1]

Off-list selling, then, is peculiarly characteristic of oligopolies which are trying to co-ordinate policies, that is practicing collusion or price leadership. The relative inflexibility of the nominal prices of

6. *Ibid.*, pp. 425–26. Cf. n. 7, above, p. 123.
7. *Ibid.*, p. 428.
8. *Op. cit.*, p. 77.
9. *Op. cit.*, p. 426.
1. *Ibid.*, pp. 425–26. For a group of non-co-ordinating oligopolists, cf. Nicholls, *op. cit.* During the period of price competition (1917–1923), "net prices of the three brands were much more nearly equal than their list prices, the differences in discounts (regular and secret) largely offsetting differences in the latter" (p. 55). In the subsequent period of "price leadership, virtual list and net price identity and increasingly non-aggressive price policy" (p. 78), Reynolds "apparently shaded the price of Camels . . . such price-shading was probably indirect and concealed through inclusion of special deals" (p. 89).

these two industry groups, which Stigler adduces, does not justify the presumption that their transaction prices were also relatively inflexible.

C. *Collusion and Price Leadership*

Stigler states that "there will be no kink when the oligopolists enter into explicit collusion"[2] and that (barometric) price leadership is "explicitly a device to insure that there will be no kink, or that the kink will not prevent readjustment of price to important changes in cost or demand conditions."[3]

I cannot quarrel with the latter statement. In fact the stated motivation applies equally to price leadership and to explicit collusion, which, indeed, is merely a "higher" form of oligopolistic co-ordination. But the fact that collusion or acquiescence to price leadership is *intended* to remove the kink by no means implies that it does so.[4] Even in the event of explicit collusion oligopolists must, in the American legal environment, always fear nonconformance of rivals. They seldom, or never, can be assured that rivals will follow a price revision, if it appears to be disadvantageous. They must also, if the revision is downward, fear the outbreak of price warfare. These uncertainties imply the unreliability of forecasts of demand based on the unkinked demand functions. If an oligopolist fears nonconformity to upward price revisions, his demand curve is in fact, or may be, still obtusely kinked. If he fears price warfare, estimates of short-run demand are irrelevant to his pricing decisions. In so far, then, as Stigler's two groups of oligopolists (those who entered into explicit collusion, and those with price leadership) were unassured, they did not base decisions on unkinked demand curves.

It is interesting, in this connection, that the period of explicit collusion of rayon producers, which Stigler cites, endured only seven months, and also that prices of pineapples changed 17 times during the 39 months preceding August 30, 1932, when six growers entered into a ten year agreement, and 8 times in the following 57 months of collusion.[5] The brevity of the period of collusion and the preva-

2. Stigler, *op. cit.*, p. 428.
3. *Ibid.*, p. 433.
4. Cf. Fritz Machlup, *The Economics of Sellers' Competition* (Baltimore: Johns Hopkins Press, 1952), p. 474, n. 30. "But explicit collusion may imply a monopolist's demand curve for the group of oligopolists as a whole . . . not for the individual oligopolist when he considers whether or not he should stick to it."
5. Stigler specifies, *op. cit.*, pp. 428–29, precise dates and numbers of price changes only for rayon and pineapples among the five instances of explicit collusion.

lence of off-list selling during collusion in the rayon industry and the number of price changes of pineapples before and after collusion indicate that the oligopolists of these two industries, even during the periods of collusion, could not, unless they were extraordinarily naive, have felt assured that rivals would conform to the agreements. In short, they probably did not base pricing decisions on estimates of the unkinked curves.

The fact of consistently depressed markets in the period of Stigler's comparisons is significant also with respect to the specific purpose of oligopolistic collusion or agreement, and the pricing patterns which, so far as compliance is assured, emerge. In depressed and declining markets, pressure on prices is predominantly downward, and the primary motivation of agreement, whether explicit or spontaneous, is less likely to be the hope of co-ordinately raising prices than the hope of preventing downward revisions and price warfare. By an agreement, the rivals may expect merely to prevent the collapse or further deterioration of the existing price structure.[6] The immediate effect of an agreement, if it succeeds, is, then, to end or forestall price warfare, or to raise prices from the unsatisfactory levels to which they had fallen. But if demand continues to decline (1931–1932), there is little likelihood of continued or even occasional upward price revision. Downward flexibility is at the same time discouraged by conjectures, which, of course, are strengthened by the fact of agreement, that rivals will meet price reductions, and by the fear that the agreement will collapse. The co-ordinating oligopolists refrain from upward price revisions either because they fear nonconformity of rivals, or because such revisions, even if general, would be unprofitable. In the former case, demand curves of the colluding oligopolists are kinked, and in the latter case, they may be unkinked. But the question is academic. Even *successful* agreement, whether explicit or spontaneous, implies not flexibility, but comparative stability of prices, so long as the pressure of the market on prices is predominantly downward.

D. *Price Rigidity of Monopolies and Heterogeneous Oligopolies*

The first and fourth of Stigler's comparisons are subject in full to only one of the preceding criticisms, viz., the first — his inclusion in his measures of flexibility of prices during all types of market situa-

6. Markham, *op. cit.*, p. 44, says: "Whatever the long-run effects of the [rayon] agreement may have been, *the immediate aim to stabilize prices* seems to have met with little success." (My italics).

tions. Let us therefore accept, for the sake of argument, the findings of these two comparisons, and try to re-evaluate them.

The first and fourth comparisons are grouped together because monopoly, for the present purpose, differs from oligopoly with heterogeneous products only in the lesser "goodness of substitutes."[7] Reduction of cross-elasticity by differentiation means reduced dependence of demand on rivals' reactions. If an oligopolist (under heterogeneous products) thinks cross-elasticity is insignificant, or chooses to ignore it, his demand curve is, as in monopoly, unkinked. His price should, as Stigler posits, be flexible compared with those of oligopolists under homogeneous products, in so far as the latter base decisions on (obtusely) kinked curves, and in so far as other influences leading to rigidity do not outweigh the effects of the kinks.

The last qualification is the nub of the matter. Stigler cites altogether five influences, aside from obtuse kinks, which make for price stability, to wit: "long-run considerations," "administrative weaknesses in collusion," "cost of price changes," "fear of governmental attention," and "conservatism that comes with size."[8]

Except for the second, all apply in more or less equal measure to oligopolies and monopolies, and to oligopolies with and without product differentiation. It is, moreover, impossible to isolate the several influences.[9] Stigler's first and fourth comparisons, so far then as they are valid, indicate that the specific kink influence may be frustrated by other influences making for price flexibility, but not that it is nonexistent. Stigler, it seems, required of the kink theory that its effect outweigh the effects of all other influences, which is too much. These comparisons may indicate — if any indication were needed — that obtuse kinks are not the sole nor even the predominant influence determining rigidity, and thus reduce the importance of the theory. But they do not prove that it is incorrect.

To summarize the argument of this section: Stigler's compari-

7. Stigler says, *op. cit.*, p. 434, that "we might, with some justification, have designated our monopolies (aluminum and nickel) as oligopolies with differentiated products." Cf. n. 6, above, p. 128.

8. *Ibid.*, pp. 418–19, 428.

9. Stigler, because the implication that the kink disappears when there are "many rivals," is identical "with that of the neoclassical theory," (*ibid.*, p. 431), makes no comparison of the flexibility of prices of oligopolistic and non-oligopolistically competitive industries. Even, however, if such comparison were to indicate the greater rigidity of the oligopolistic prices, the specific influence of the kinked curves would again not be shown in isolation, because other influences which affect price behavior among oligopolists are absent or less powerful in non-oligopolistically competitive markets.

sons, because they are based on changes of nominal or list prices, and because periods when oligopolists determined prices on the basis of kinked curves are not segregated, are inconclusive with respect to the rigidity of transaction prices implied by the theory. Two of the comparisons are faulty also because they depend on the assumption that price policy co-ordination will eliminate the kink, whereas, although such is indeed the motivation of the co-ordinating oligopolists, it is but rarely achieved. Finally, the apparent exceptional rigidity of monopoly prices and prices of heterogeneous oligopolies may indicate only the force of those influences other than the kinked demand curves which make for price rigidity.

III. Conclusions

1. Stigler thinks that the price histories which he relates indicate "only that not every oligopoly has reason to believe that it has a kinky demand curve," and this, as he says, "most adherents of the theory would readily concede."[1] But the experience that rivals, after demand has sharply risen, will follow a price increase, does not contradict the validity of the possible previous and subsequent belief in the kink.

2. Since obtuse kinks imply comparative rigidity of prices, it is disappointing to find that Stigler's first and fourth comparisons, whatever their shortcomings, do not support the implication. They indicate, however, not the incorrectness of the kinked curve construction, but, at most, that the influence of obtuse kinks is not the sole, and perhaps not even the most significant, among the various influences which reduce price flexibility.

3. Since the validity of the conjectures which determine the kink is likely to be revealed by experience only in those exceptional cases where oligopolists have mistakenly disbelieved, and since the rigidity effect of obtusely kinked curves is, apparently, often submerged by other influences affecting flexibility, the validity of the construction cannot, unfortunately, be corroborated — or proven — by price histories or general comparisons of the types adduced by Stigler. Adherents of the theory, unless they can segregate periods when oligopolists actually determined prices on the basis of obtusely kinked demand curves, are thrown back on the insecure foundations of a priori reasoning which plagues all marginal analysis and on inferential analysis of the motivations of entrepreneurial decisions. Evidence,

1. *Ibid.*, p. 425.

then, of oligopolistic belief in the kink, and of its influence on oligopolistic behavior must be sought in industry studies, such as Nicholls' and Markham's studies of the tobacco and rayon industries, wherein the histories of transaction as distinguished from list prices, and the specific motivations of oligopolists' decisions are analyzed in detail.

C. W. EFROYMSON.

BUTLER UNIVERSITY,
INDIANAPOLIS, INDIANA.

THE TABLEAU ÉCONOMIQUE AS A
SIMPLE LEONTIEF MODEL

By Almarin Phillips

I. Introduction

The primary purpose of this note is to provide an exposition of François Quesnay's *Tableau Économique* as a simple Leontief input-output system. Presented in this way, the *Tableau* is much easier to understand and Quesnay's contribution to economic thinking will perhaps be more apparent. Secondarily, the exposition may serve as an introduction to input-output economics for those not yet acquainted with that technique.

The *Tableau* was first printed in 1758. No copies of the first edition are known to exist but several of the later editions are readily available. The edition most frequently reprinted is that of 1759, which was discovered by Dr. Stephan Bauer and published by the British Economic Association in 1894.

Recorded opinions of the *Tableau,* even those by more recent students of economic doctrines, differ widely. Mirabeau declared it to be one of the three greatest discoveries since the world began.[1] Karl Marx extended himself almost as far when, writing of the *Tableau,* he opined that, "Never before had thinking in political economy reached such heights of genius."[2] And one need not search far to find words of praise from modern writers. Newman thinks of the *Tableau* as "giving [economics] its start as a science" and as "the earliest successful attempt to analyze a nation's wealth on a macro-economic basis."[3] Essentially the same view can be ascribed

---

1. Marquis de Mirabeau, *Philosophie Rurale ou économie générale et politique de l'agriculture, pour servir à l'ami des hommes* (Amsterdam, 1766), tom. i., pp. 52, 53 (Quoted by Adam Smith, *The Wealth of Nations,* Cannan ed., II, 177 n.).
   2. Karl Marx, *A History of Economic Theories,* ed. Karl Kautsky and trans. Terence McCarthy (New York: The Langland Press, 1952), pp. 67, 68.
   3. P. C. Newman, *The Development of Economic Thought* (New York, 1952), pp. 34 ff.

to Haney[4] and to Roll.[5]   Gide would agree that the physiocratic school, by attempting to define a natural order, "marks the beginning of . . . the science of Political Economy,"[6] but in his opinion the enthusiasm of Mirabeau over the *Tableau* "is almost incredible."[7] Speaking further of Mirabeau and others among Quesnay's admirers, Gide feels that, "The rediscovery of the millions intoxicated them, but, like other mathematical economists of today, they forgot that at the end of their calculations they only had what they had assumed at the beginning.   It is evident that the table proves nothing as to the essential point in their system, namely, whether there really exists a productive and a sterile class."[8]

A caustic critic, Gray, writes that the *Tableau* should perhaps be "reduced to an embarrassed footnote"[9] in the history of economic thought.   "It may be doubted whether it will ever be anything but a vast mystification, a subject to be treated gingerly by commentators, rendered uneasy by the feeling that they do not quite understand what they are talking about."[1]

However vast the mystification, Leontief introduces his work by saying that it "may best be defined as an attempt to construct . . . a *Tableau Économique* of the United States,"[2] explicitly recognizing the similarity between his method and that of Quesnay.

II. Quesnay's Tableau

Quesnay presented his *Tableau* in two forms.   The more familiar one is from the 1759 edition discovered by Dr. Bauer.[3]   This table is not, however, an analysis of the whole of an economy but rather a diagrammatic representation of the regeneration of income due to the expenditure of a member of the proprietor class.   It illustrates how 600 livres received by one landowner as rent are spent, half on agri-

4. L. H. Haney, *History of Economic Thought* (4th ed.; New York, 1949) pp. 169 ff.

5. E. Roll, *A History of Economic Thought* (New York, 1942), pp. 132 ff.

6. C. Gide and C. Rist, *A History of Economic Doctrines*, 2d English ed., p. 22.

7. *Ibid.*, p. 37.

8. *Ibid.*, p. 40.

9. A. Gray, *The Development of Economic Doctrine* (London, 1931), p. 106.

1. *Loc. cit.*

2. W. W. Leontief, *The Structure of American Economy: 1919–1939* (2d ed.; New York, 1951), p. 9.   For similar comments, see W. Duane Evans and Marvin Hoffenberg, *The Interindustry Relations Study for 1947*, Bureau of Labor Statistics, 1951, p. I-1 and A. Charnes, W. W. Cooper and A. Henderson, *An Introduction to Linear Programming* (New York, 1953), p. 2.

3. This table is so well known that it is not reproduced here.   See Gray, *op. cit.*, facing p. 106, or Newman, *op. cit.*, p. 36 for reproductions.

cultural products and half on the products of artisans. Then, in their turn, farmers buy industrial products, artisans buy food and raw materials, and so on back and forth down the table in ever lessening subdivisions of the original 600 livres spent. According to Quesnay's interpretation, with a given amount of permanent land improvements and fixed capital (*avances foncières* and *avances primitives*), the value of production in agriculture exceeds the cost of labor, seed, other recurring annual costs, interest and depreciation (*avances annuelles*) and, with the respending of the original 600 livres going on indefinitely, this surplus over cost will result in a total of 600 livres being repaid to the proprietor. In fact, of course, the 600 livres go to the proprietor class, not necessarily to the proprietor who initiated the expenditure stream.

This "zigzag" table is difficult to interpret. Later editions of the *Tableau* simplify the illustration by changing the figures to those of proprietors, farmers and artisans, each as a whole class, and by giving only their total annual receipts and expenditures. Table I is reproduced from Quesnay's "Analyse du Tableau Économique" which was published in the *Journal d'Agriculture, du Commerce et des Finances;* June, 1766.[4] In this version, proprietors spent two milliards (two thousand million livres), one milliard on agricultural products and one milliard for the goods of artisans. These are shown in Table I by the two dotted lines from the proprietors' two milliards of revenue down to the farmers (*la classe productive*) and artisans (*la classe sterile*) respectively. Artisans spend two milliards also, but all of this is for agricultural products — food and raw materials. The table shows two dotted lines from them to farmers to represent these purchases. Farmers purchase only one milliard of goods, this from the artisans, but they retain two milliards of their own production. Quesnay's total reproduction of five milliards is the sum of agricultural production and is composed of the one milliard purchased by proprietors, the two milliards purchased by artisans and the two milliards retained by farmers. This is the clearest diagram of the *Tableau* that Quesnay provided.[5] In the next section the same data will be presented as an input-output model and the nature of the Quesnay analysis becomes more clear.

4. Reprinted in Eugene Daire (ed.), *Les Physiocrats*, Vol. I (Paris, 1846), p. 65.
    5. A crucial element not shown by the lines in the formal table is the "purchase" by the farmers of two milliards of "rental services" from the proprietors, with which these latter begin the following year.

TABLE I

<small>FORMULE DU TABLEAU ÉCONOMIQUE
REPRODUCTION TOTALE: 5 MILLIARDS</small>

	Avances annuelles de la classe productive	Revenu pour les propriétaires des terres, le Souverain et les décimateurs	Avances de la classe sterile
	2 milliards	2 milliards	1 milliard
Sommes qui servent à payer le revenu et les intérêts des avances primitives	1 milliard 1 milliard 1 milliard		1 milliard 1 milliard
			Total, 2 milliards
Dépenses des avances annuelles	2 milliards		dont la moitié est retenue par cette class pour les avances de l'année suivante

Total, 5 milliards

III. The Tableau Économique as an Input-Output System

A three industry closed Leontief model is of the form,

$$
\begin{aligned}
&+ (1 - a_{11})X_1 && - a_{12} X_2 && - a_{13} X_3 = 0 \\
&- a_{21} X_1 && + (1 - a_{22})X_2 && - a_{23} X_3 = 0 \\
&- a_{31} X_1 && - a_{32} X_2 && + (1 - a_{33})X_3 = 0
\end{aligned}
$$

(1)

where X_1, X_2, and X_3 are the rates of output of each of the three industries and the a_{ij} coefficients represent the requirements of the jth industry for the ith industry's product per unit output of the jth industry. That is, for example,

$$
a_{12} = \frac{x_{12}}{X_2}
$$

where the x_{12} is the amount of the first industry's product which the second industry purchases. These equations describe the truistic condition that, if the three industries comprise the total economy, the total output of an industry is equal to the amount retained by that industry plus the amount sold to the other industries. More formally, the equations state that the amount of the product of an industry

which it does not retain, $(1 - a_{ii})X_i$, less the amounts purchased from it by other industries is equal to zero.

From the *Tableau*, let Farmers be Industry I, Proprietors be Industry II, and Artisans be Industry III. We may then prepare a transactions table from which to compute the required coefficients from the information which Quesnay provided.

TABLE II

TRANSACTIONS TABLE FOR THE TABLEAU ÉCONOMIQUE

(Value of real goods in milliards)

Producing Industry	Purchasing Industry			
	I Farmers	II Proprietors	III Artisans	Total Production
I Farmers	2	1	2	5
II Proprietors	2	0	0	2
III Artisans	1	1	0	2
Total Purchases	5	2	2	9

Farmers produce a total of five milliards, two of which they keep. One milliard is sold to proprietors and another two milliards is sold to artisans. Farmers purchase the two retained milliards of their own goods, two milliards of rental services from proprietors and one milliard of goods from artisans. Proprietors "produce" two milliards of rental services, all of which is sold to farmers. Artisans produce two milliards of goods, half of which is purchased by proprietors and half by farmers. The whole *Tableau* is thus simply presented by this two-way transactions diagram.

Using the values of Table II, the input-output coefficients of the *Tableau* appear as follows:

$$a_{11} = .4 \qquad a_{12} = .5 \qquad a_{13} = 1.0$$
$$a_{21} = .4 \qquad a_{22} = 0 \qquad a_{23} = 0$$
$$a_{31} = .2 \qquad a_{32} = .5 \qquad a_{33} = 0$$

Substituting these values into the closed Leontief system of equation set (1), we have,

$$
\begin{aligned}
+ .6\,(5) - .5\,(2) - 1.0\,(2) &= 0 \\
- .4\,(5) + 1.0\,(2) - 0\ \ (2) &= 0 \\
- .2\,(5) - .5\,(2) + 1.0\,(2) &= 0
\end{aligned}
$$

(2)

The actual values for X_1, X_2 and X_3 which Quesnay gave are only one of an infinite number of values for them which would satisfy the Leontief equations. Being a closed system of linear and homogeneous equations and with no part of the demand for products given autonomously, the equations are met by any set of values for X_1, X_2 and X_3 which are in the same proportions as 5, 2 and 2 respectively. It is impossible, given only the coefficients, to solve for anything other than the proportion which the output of one "industry" is to the output of one of the others. On the other hand, given the coefficients and the rate of output for any one industry, all the other rates of output are determined.

IV. Suggested Interpretations of the Tableau Économique

The presentation of the *Tableau* in input-output form makes it possible to clarify some of the more obscure parts of Quesnay's analysis. Quesnay was apparently troubled by the problem of explaining where the money came from to get his process started. To solve this he used a sort of period analysis not completely unlike that used in the Robertsonian method for analyzing national income. The two milliards which proprietors spend are described by Quesnay as being those received from the previous year's rent. That is, spending in one income period is the result of income received during the previous period. Similarly, of the two milliards received by artisans, half is retained for initial expenditures in the following year.

There is surely no error in putting such leads and lags into the *Tableau*, but they do make it more difficult to understand. Table II indicates that Quesnay's economy could more easily be pictured as a circular flow, since each of the industries has total receipts (production) equal to total expenditures (purchases).[6] No leads and lags are necessary, each sector simply spending in each income period the receipts of the same period.

Presenting the *Tableau* as an input-output system also clarifies the frequently mystifying nature of the *avances annuelles* and the

6. In agriculture, as already noted, purchases include goods produced by farmers and retained by them for their own use. These may or may not involve monetary transactions but, in either case, need not upset the equality between income and expenditures.

avances primitives.   In the edition of the *Tableau* from which the figures for Table II are taken, Quesnay assumed that fixed capital in agriculture, excluding land and permanent improvements to land (*avances foncières*), amounted to ten milliards.   He called this *avances primitives* and there is some confusion regarding its maintenance. At one point Quesnay speaks of an interest payment amounting to 10 per cent of the ten milliards,[7] but this is not shown in the *Tableau*. At another point, speaking of the same milliard of expenditures by farmers, Quesnay describes them as purchases from the artisan class,[8] which are shown in the *Tableau*.   It seems preferable to regard the 10 per cent rate as one of depreciation rather than interest.   Then 10 per cent of the fixed capital wears out each year but is replaced by purchases from the artisans.   In this way fixed capital is assumed to remain constant and the stationary, circular equilibrium is maintained.   Such an interpretation seems to capture Quesnay's real meaning.[9]

The *avances annuelles* are most simply described as the non-capital, nonrent costs of production and are assumed by Quesnay to amount to two milliards per year.   In the *Tableau* these are represented by the two milliards of their own production which farmers retain.   These are the costs for seed, fertilizer, agricultural wages (subsistence) and other recurring annual expenses other than depreciation.

The total of the *avances annuelles* and the *avances primitives* Quesnay calls *les reprises*, all production costs other than rent.   They amount, of course, to three milliards.   In Table II, *les reprises* are shown as the two milliards which farmers purchase from themselves and the one milliard which they purchase from artisans.

The physiocratic "net product" is equivalent to agricultural rent payments.   And these bear a striking resemblance to the then unknown Ricardian differential surplus, the difference between the value of the product of the land and the cost of nonland productive factors.   Only

7. Daire, *op. cit.*, p. 62, "*l'autre milliard* est prélevé par cette même classe sur ses ventes pour les intérêts des avances de son établissement" (italics in original).

8. *Ibid.*, p. 60, "en achats d'*un milliard* d'ouvrages qu'elle paye à la *classe sterile*" (italics in original).

9. Compare this interpretation with Marx, *op. cit.*, p. 65: "In his example, Quesnay evaluates this fixed capital at ten billions, the interest on which, at ten per cent, would amount to one billion.   By this he does not mean surplus value but merely the cost of repair and upkeep of the buildings, restocking herds, setting aside of a reserve of funds to compensate for occasional losses caused by crop failures, epidemics, floods, and to improve the soil or to extend the tillage. The third billion, therefore, corresponds in amount to that part of the fixed capital which is equal to the wear and tear and which reappears in the values of the annual product.   It is mainly consumed in the agriculture which produces it."

agriculture appears to produce a net product since, by assumption, only farmers pay rent. Perhaps the typically small-scale craft shops of artisans, principally owner-operated, concealed from the physiocrats the payment of rents to nonproductive property owners in other than agricultural production. Certainly it would have been illogical for the physiocratic doctrine to accept the agricultural net product theory had rent payments been assumed for the other industries.[1]

V. Conclusions

Quesnay's *Tableau Économique* may be viewed as a static, closed Leontief system. Presented in this way, the circularity of the economy becomes more apparent and some interpretive difficulties are lessened. More important, Quesnay's brilliance in formulating a system of general interdependence becomes clear by such analysis. It leads one to consider his contribution to economic thought much as did Marx: "Never before had thinking in political economy reached such heights of genius." The *Tableau Économique* marks the beginning of general equilibrium economic theories and well deserves mention in preface to Walras, Pareto, Cassel and Leontief.

Almarin Phillips.

University of Pennsylvania

1. Compare this with G. Weulersse, *Encyclopaedia of the Social Sciences*, 1931 edition, Vol. V, p. 349, "Trade and industry appeared sterile because the physiocrats could not discover that they yielded a regular surplus profit similar to the agricultural net product."

THE EXTENT AND TRENDS OF RAIDING
AMONG AMERICAN UNIONS

By Joseph Krislov[*]

I. Introduction

Discussions of the problem of raiding among American unions have taken place without any grounding in statistical fact. This note proposes to lay the foundations for a statistical examination of the problems of raiding.[1]

A raid is defined as an attempt by one union to obtain bargaining rights in a unit in which another union holds certification from the National Labor Relations Board or has been given an exclusive bargaining contract by the employer. The definition is focused on the attempt itself, regardless of the result of the election. Nevertheless, only if an election were held in the unit, would it appear in this statistical analysis.[2] The petitioning union may withdraw its petition, and the NLRB may dismiss petitions for a number of reasons. In 1951, for example, only 62.5 per cent of the representation cases closed actually culminated in an election.[3] Some of the units in which an election was not held were organized, and there was, in a sense, a raid. In addition, a union attempting to raid may be unable to secure the 30 per cent showing necessary to support a petition. Therefore, the union would not petition for an election, and the case

---

1. See my previous companion article, "Raiding Among the 'Legitimate' Unions," *Industrial and Labor Relations Review*, Vol. VIII, No. I (Oct. 1954), pp. 19–30.

2. A limitation of this article is the possible existence of raiding in bargaining relationships which do not stem from NLRB certification. Particularly important are the relationships in the railroads, air lines, and the maritime industry; and the relationships that stem from certifications by state labor boards. Furthermore, there exist bargaining relationships in which recourse to any public board's processes is rare. Recognition in these industries is obtained by other methods. It is doubtful, however, whether a study of these areas would invalidate our general conclusions. See my "Raiding and Rivalry Among American Unions: A Statistical Analysis." Unpublished Ph.D. thesis, University of Wisconsin, 1954, pp. 65–66.

3. National Labor Relations Board, *16th Annual Report*, 1951, pp. 291, 304.

would not appear in NLRB records. This analysis, therefore, undoubtedly understates the number of attempted raids.[4]

The basic source of data is the NLRB files of election results. Seven years have been chosen for study, each representing a distinct period in the organizational activities of American labor unions. The fiscal year 1940 was selected as representative of the prewar period, while fiscal 1943 portrays the war period. The fiscal year 1946 depicts the pre-Taft-Hartley postwar period of 1945–47. The post-Taft-Hartley years of 1949 and 1950 reveal organizational activities during the period when many unions did not comply with the non-Communist affidavit provisions of the Act. By the end of the fiscal year 1950, most unions had complied with the filing requirements. A study of 1951 and 1952 was made to represent the period before American unions became interested in no-raiding agreements.

Within each year, the second six months of the fiscal year (the first six months of the calendar year) have been selected for intensive study. The elections in which two unions appeared on the ballot in a previously organized unit constitute a portion of the units raided. An additional source of raiding is the election contest in which the recognized union does not appear on the ballot. In these contests, the raiding union captures the unit without a fight.

Two counts of raiding can be made. The first involves a count of individual elections, disregarding the numbers of workers involved. The second considers only the workers involved, disregarding the number of elections. Although the mechanics of both counts are similar,[5] the results diverge somewhat.

4. Changes in union affiliations which are made without an NLRB election should not be construed as falling within the confines of this study. No accurate record is available regarding the number of these "switches." Moreover, the change in union affiliation is often made without a secret ballot being taken. While the "switch" may actually reflect employee wishes, it need not necessarily do so.

5. In each year approximately half of the multi-union elections have been studied. A percentage is obtained by dividing the totals in the elections studied by the published results of all multi-union elections conducted by the Board. The information obtained from the sample cases is assumed to be representative and is then projected for the entire year. The result is taken as the measure of raiding in the multi-union classification of representational elections. To ascertain the extent of raiding in the one-union category, a less direct procedure had to be employed. The crucial question which must be answered before any computation can be made is this: What percentage of the universe of "one-union raids" does the sample represent? An assumption is made that the percentage of cases studied or workers involved in the one-union contest is the same as that in the multi-union category. On this assumption, the task is one merely of projecting the one-union results. The results from both types of elections are then added. These totals are divided into the total number of elections conducted or voters involved in the year in question to ascertain the over-all extent of raiding. Year by year comparisons will reveal the presence or absence of any trend.

Certain unions can be considered as falling outside the circle of "bona-fide" trade unions, recognized as such by the major segments of the labor movement. Organizations falling within this category include local independent unions, whose membership is limited to one company, and the left-wing unions expelled from the CIO.[6] Trade unionists regard all local independents as company dominated unions. It is untrue, however, that all such unions are dominated by employers.[7] The exclusion of the left-wing unions is made to conform to the announced policy of the two labor federations, which maintains that the leaders of these unions are converting them into a political agency to serve a foreign power. Public policy also seems to be directed toward crippling the left-wing unions.[8]

In accord with these considerations, raids were broken down to determine the union that was raided. Tables are presented indicating the type of union that was raided. The first column of Tables I and II takes account of all raids against all certified or recognized unions. The second column omits raids against left-wing unions, which began in 1949; and the third omits raids against local independents as well as those against left-wing unions. This last column may be viewed as giving the number of raids involving two "bona-fide"[9] unions, i.e., those unions affiliated with the AFL or CIO, and the national independents. However, a raid against a "bona-fide" union may occasionally stem from a left-wing union or from a local independent.

Tables I and II indicate the number and the percentage distribution of raids compared with all elections, broken down into three classifications.

6. The eleven unions expelled were: the United Electrical Workers; the Farm Equipment Workers; the United Public Workers; the United Office and Professional Workers; the Mine, Mill and Swelter Workers; the Food Tobacco and Agriculture Workers; the American Communications Association; the United Fur and Leather Workers; the International Longshoreman's and Warehouseman's Union; the Marine Cooks and Stewards Union; and the Fisherman and Allied Workers.

7. The most extensive study of independent unions was made by Mary Klemn, "Some Rivals of the AFL and CIO: A Study of Unaffiliated Unions." Unpublished Ph.D. Thesis, University of Wisconsin, 1943. Her investigations are summarized in "The Rise of Independent Unions and the Decline of Labor Oligopoly," *American Economic Review*, XXIV (Sept. 1944), 76–86. She writes that "for the most . . . they [the independents] are genuinely independent of the employer." p. 76. Cf. her comments on p. 80.

8. *Business Week*, Nov. 7, 1953, p. 166. Public policy toward the left-wing unions was finally embodied in statutory law, the Subversive Activities Control Act of 1950, 83d Congress, 2d Session.

9. Merely to facilitate expression, we shall use the term "bona-fide" unions rather than the more accurate but cumbersome phrase "affiliated unions and national independents."

II. Number of Elections

TABLE I

Number of Raids: Total and by Type of Union

Year	All Elections	Elections Involving Raids	Raids Excluding Those On Left-Wing Unions	Raids Excluding Those On Left-Wing and Local Independents
1940	1192	78	—	62
1943	4153	490	—	288
1946	5589	612	—	422
1949	5514	814	674	576
1950	5619	1,384	736	612
1951	6432	959	757	578
1952	6766	1,186	1,045	881

Source: National Labor Relations Board.

TABLE II

Raids as a Percentage of All Elections

Year	Percentage of Raids All Elections	Percentage of Raids Excluding Those On Left-Wing Unions	Percentage of Raids Excluding Those On Left-Wing and Local Independents
1940	6.54	—	5.20
1943	11.80	—	6.93
1946	10.95	—	7.55
1949	14.76	12.22	10.45
1950	24.45	12.81	10.71
1951	14.91	11.77	8.98
1952	17.53	15.44	13.02

Source: National Labor Relations Board.

It is evident that the number of elections involving raids has never been large compared with the total number of elections. Considering only the absolute number of raids, there appears to be a persistent, upward trend. Part of this increase can be attributed to the gradual increase in the number of elections held each year. However, there is a slight upward trend in the percentage of raids each year.

Raids against left-wing unions have been extensive. Particularly significant were the raids against the left-wing unions in the year of their expulsion from the CIO, i.e., fiscal 1950. The bulk of these contests were between the independent United Electrical Workers, and the newly formed International Union of Electrical Workers of

the CIO. There is much doubt whether these contests should be considered raids. Many of the locals in these contests were merely deciding with which faction of the old "United Electrical Workers" they wished to affiliate themselves. In a number of these contests, however, the AFL Electrical Workers and the Machinists joined the new CIO union in attacking the independent United Electrical Workers. Even if no third union appeared, it does not seem wise to treat these contests as attempts at new organizing or to create a special category for them.

Exclusive of the attacks on left-wing unions, there is a steadily increasing number of raids. The percentage of raids involving both "bona-fide" unions and local independents has fluctuated somewhat. There appears to be a noticeable increase from 1940 to the years immediately preceding the Taft-Hartley Act, and a continued but smaller increase since the passage of the Act. The percentage increase in 1952 is substantial, and could possibly have been the forerunner of a continuous yearly increase in raiding.

Raids on local independents fluctuated considerably during the years studied. The continuance of these raids is indicative of the unwillingness of the labor movement to accept these groups as "bona-fide" unions. If these raids are successful, it is an indication that local independents are fruitful sources from which the "bona-fide" organizations may extend their membership.

What has been the development of attacks on the "bona-fide" unions? Except for 1951, there has been a slight, persistent increase yearly in both absolute numbers and percentages. An explanation for the slight decline in 1951 may lie in the negotiation of long-term contracts previous to that year. The Board's willingness to honor long-term contracts closed these units to elections. As a result, many unions contemplating raiding may have been prevented from petitioning and going to an election. In 1952 raiding between the "bona-fide" unions increased substantially, and exceeded any previous year studied.

III. Voters Involved

Tables III and IV present the number of eligible voters and the percentage distribution of the voters in elections involving raids, compared with the total number of eligible voters.[1] The same statis-

1. Two qualifications should be noted as to the data relating to the number of eligible voters in the elections studied. It is impossible to make an accurate estimate of the number of voters involved in each election category for 1940. Consequently, that year is eliminated from this part of the analysis. Moreover, it has been assumed throughout that the period studied was entirely representative of the universe. While this can be asserted confidently with respect to the

tical breakdown involving local independents, left-wing unions, and "bona-fide" unions is presented.

TABLE III

Eligible Voters in Raids

(Thousands of Voters)

Year	Voters in All Elections	Voters in All Elections Involving Raids	Voters in All Raids Excluding Those On Left-Wing Unions	Voters in All Raids Excluding Those On Left-Wing and Local Independents
1943	1,402	319.7	—	188.7
1946	846	281.9	—	175.6
1949	589	244.6	155.6	99.5
1950	840	561.7	234.5	195.5
1951	666	284.3	203.6	132.2
1952	771	356.5	304.4	249.7

Source: National Labor Relations Board.

TABLE IV

Percentage of Eligible Voters Involved in Elections Growing Out of Raids

Year	Percentage in All Raids	Percentage in All Raids Excluding Those On Left-Wing Unions	Percentage in All Raids Excluding Those On Left-Wing Unions and Local Independents
1943	22.85	—	13.46
1946	33.32	—	20.75
1949	41.53	26.42	16.89
1950	66.52	27.68	23.04
1951	42.69	30.57	19.85
1952	46.24	39.48	32.39

Source: National Labor Relations Board.

Perhaps the most important aspect shown in Table IV is that a much larger percentage of voters are involved in elections growing out of raids than would be suggested by Table II. The inference to be drawn is simply that raids are most likely to occur in larger plants.

number of elections, it is open to question in relation to the number of voters. In any given year, it may happen that the Board conducts relatively few elections in extremely large units. In 1943, for example, only 6 per cent of the contests were conducted in units with over a thousand employees. There are many ways in which these few elections conducted in any given year might be arrayed so as to make our samples unrepresentative, and short of studying all the cases, there is little that can be done to correct this possible weakness. However, the aggregative results of the study do not indicate any distortion.

This is to be expected, for there are few large plants still unorganized. New organization, therefore, takes place primarily in small or medium-sized plants. And with the exception of craft groups seeking to sever from industrial units, the larger the plant the more likely it is to invite raids.[2]

The number of workers involved in all raids seems to indicate that raiding is more widespread than is revealed by the number of elections growing out of raids. Because of significant variations in the absolute number of voters involved, it is desirable to focus attention primarily on the percentage table. There appears to be a persistent upward trend in the percentage of voters involved in all raids. The year 1950 stands out as a banner year, but about one-half of all eligible voters in the raids of that year were in elections involving left-wing unions.

Exclusive of raids against left-wing unions, there appears to be an increase in raiding from the war period, represented by 1943, to the postwar pre-Taft-Hartley period, depicted by 1946. The post-Taft-Hartley years show a slight decrease in the percentage of voters involved in raids. This resulted from increased craft-raiding, following the passage of the Taft-Hartley Act.[3] It is doubtful whether the specific sanction of craft severance in the Act changed the law on the subject.[4] Across the board wage increases had continually narrowed the wage differential between skilled and unskilled workers. Many craft employees were undoubtedly dissatisfied and the Taft-Hartley provision served to dramatize their plight. As a result, the craftsmen became more conscious of the possibilities of severance after the passage of the Act. However, the increased possibility for craft severance did not deter industrial raiding. In 1952 the percentage of workers involved in raids considerably exceeded the pre-Taft-Hartley percentage.

As with the number of elections, the number of workers involved in raids against local independents has fluctuated considerably. The number of such elections is small, but the number of voters involved is substantial. This fact would indicate that the units are large, and represent the last strongholds of what might be termed sentiment for "local unionism."

The raiding of "bona-fide" unions has fluctuated somewhat.

2. Organizers of several unions indicated that a discontented nucleus could always be found in large plants. In addition, the prize of the large plant was exceedingly tempting.

3. Krislov, "Raiding Among the 'Legitimate' Unions," *op. cit.*, p. 20.

4. See my "Raiding and Rivalry Among American Unions: A Statistical Analysis," *op. cit.*, pp. 40–42.

There was an increase in the postwar period, and a leveling off in the percentages during the years immediately following the passage of the Taft-Hartley Act. The increase in voters involved in raids in 1952 is striking, as it is coupled with an increased number of raids. This would tend to confirm the hypothesis that the increased possibility for severance of smaller groups did not serve to reduce the number of raids for plant-wide units.

IV. Conclusions

This study definitely establishes that raiding has been widespread and has tended to increase during the time period studied. Approximately one out of every six elections in 1952 was held in a previously organized unit. Similarly, 46 per cent of the eligible voters in election contests in 1952 cast their ballots in organized units. Surprising as these figures may be, they indicate that the labor movement did not cease to organize the unorganized. The extent to which two of the provisions of the Taft-Hartley Act are responsible for this increased raiding is difficult to ascertain. Craft raiding and attacks upon the left-wing unions probably were stimulated by Section 9*b*2 and the non-Communist affidavit requirement. Nevertheless, there might have been as many raids in both areas even without the Act. The principle of craft severance was established by the Board in November, 1946[5] — approximately nine months before the Act. Similarly, attacks on some of the left-wing unions preceded the Act[6], and continued even after the left-wing unions had complied with the non-Communist affidavit provision. The many no-raiding agreements negotiated after 1952 have probably served to decrease the number of raids. The extent to which raiding has been decreased, however, can be ascertained only by further study.

<div align="right">JOSEPH KRISLOV.</div>

OHIO STATE UNIVERSITY

5. *International Chemical and Minerals Corporation*, 71 NLRB 878 (1946).
6. The CIO Auto Workers began raiding the Farm Equipment Workers as early as 1945. *F E News*, February 1945. p. 1.

THE BALANCED BUDGET: COMMENT

By Herbert Stein

I should like to comment on a few points in Mr. Jesse Burkhead's article on "The Balanced Budget" which appeared in the May 1954 issue of this *Journal*.

1. In note 9 on page 211 of this article Mr. Burkhead says:

> "The Committee for Economic Development has proposed a budgetary policy which would set tax rates at a point to provide a surplus for debt retirement at an agreed high level of employment. These rates would be unchanged except in a severe depression or critical inflation. (*The Stabilizing Budget Policy*, New York, 1950.) This proposal, if adopted, would serve to reduce, not increase flexibility in the use of the budget as a stabilization measure, since variations in expenditures for stabilization purposes would be almost precluded."

I am afraid that these words may leave the reader with an incorrect impression of the CED proposals. In particular, the note in its general context is open to the interpretation that Mr. Burkhead is comparing the CED proposals unfavorably with an "annually-balanced budget" policy.

Now it is certainly not true that adoption of the CED budget proposals, as described in many statements since 1947, would reduce flexibility as compared with the policy of annual budget balance.

(a) Unlike the annual balance policy, the CED proposals do not call for offsetting the automatic variations in tax collections and expenditures, and the consequent variations in deficit or surplus, that would accompany variations in the national income. The CED literature regards these automatic variations as having a desirable stabilizing effect, and expects the variations to be large.

(b) CED has recommended steps to increase the automatic, stabilizing variability of both tax collections and expenditures.

(c) CED recommends the deliberate, stabilizing variation of both tax rates and expenditures to deal with serious depressions and inflations. The Committee places main reliance on variation of tax rates, chiefly because it believes that larger effects can be obtained more quickly on the revenue side of the budget than on the expenditure side.

(d) CED has recommended steps beyond strengthening automatic flexibility to increase the potential usefulness of tax and expenditure policies as stabilization measures.

153

CED's attitude toward the variation of expenditures as an anti-depression instrument was expressed in a recent policy statement.

"In general, the expansion of public works or other specific expenditures should be governed by:

(a) what can be done efficiently and flexibly,

(b) the character of the adjustment probably needed in the particular industries upon which the expenditure would have its main impact, and

(c) the value of the public benefits that would be provided by the expenditures, which must of course be appraised relative to the value of tax reduction. The reduction of taxes should be governed by what is needed in combination with an increase of expenditures so determined, and with other measures, to maintain or restore high employment."[1]

These proposals seem to me to provide a high degree of budgetary flexibility and to suggest reasonable standards for combining expenditure variations and tax variations. Even critics of the proposals have been willing to agree that they embody what Alvin Hansen called "an important advance in fiscal thinking compared with the tenets of so-called sound finance."

Undoubtedly policies are conceivable that would permit greater or more frequent variation of tax rates and expenditures than the CED proposals. But increased variability does not yield "increased flexibility in the use of the budget as a stabilization measure" unless there is reasonable confidence that the timing and amount of the additional variations will be such as to promote stability. In view of the uncertainty of economic forecasting, especially in the political atmosphere where decisions are made, and the lags in the application of budgetary changes, it is not self-evident that additional variations will promote stability rather than instability. This is the main consideration that has led CED to try to define, in general terms, the conditions in which deliberate budgetary variations are, or are not, likely to be stabilizing.

2. Going beyond the reference to the CED, I would question the basic thesis of Mr. Burkhead's paper. He seems to believe that "we continue to be dominated by adherence to the goal of a balanced budget" (p. 212), and sets out to explain what he calls "the remarkable persistence of the notion that government budgets ought to be balanced, even balanced annually" (p. 191).

Undoubtedly the notion persists. But I believe it is clear from

1. *Defense Against Recession*, Committee for Economic Development (New York, March 1954), p. 40.

the policies and statements of both the Truman administration and the Eisenhower administration that we are not *dominated* by the balanced budget idea. What seems to me important is the fact that Mr. Burkhead himself recognizes when he says (p. 215), "Nevertheless, and in spite of the difficulties, it can hardly be doubted that any modern and responsive government will use the weapons at its disposal, including budgetary policy, for purposes of economic stabilization."

3. In conclusion Mr. Burkhead summons the "critics of conservative fiscal policy" to this "remaining and challenging task": "to provide a guide to governmental budgeting which will serve two purposes simultaneously. The first purpose is to guide the transfer of resources from the private to the public sector; the second is to guide the selection of the aggregate level of receipts and expenditures with a view to stabilization. The guidelines must be widely understood and capable of political implementation."

I am sure that conservatives in fiscal policy, like the CED, who have been wrestling with this problem of synthesis for many years, will welcome the assistance of their critics. It is precisely recognition of this many-sided nature of the problem that has given conservative proposals on fiscal policy in recent years most of the features that their critics have found least satisfactory.

HERBERT STEIN.

COMMITTEE FOR ECONOMIC DEVELOPMENT
WASHINGTON 6, D. C.

REPLY

By JESSE BURKHEAD

1. Mr. Stein is indeed correct in pointing out that note 9 on page 211 of my article on "The Balanced Budget" may be subject to an incorrect interpretation. It was intended to compare CED's stabilizing budget policy, not with the practice of annual budget balance, but with what might be accomplished with a maximum use of the budget for stabilization purposes. As the article indicated, Swedish practice shows the potentialities. A thoroughgoing economic character classification of federal government activities, à la J. R. Hicks, would provide the data-base necessary for this kind of stabilization policy.

CED's original proposal rested heavily on built-in stabilizers. What can be accomplished with their use is obviously less than what can be accomplished where tax *rates* and expenditure *programs* are also variable. The CED proposal is less restrictive than annually balanced budgets but still lacks flexibility. Its limitations have been explored by Samuelson in the reference cited (*Money, Trade, and Economic Growth*), and by Hagen in an article some years ago ("Some Implications of the CED's 'Stabilizing Budget Policy,' " *Proceedings of the National Tax Association*, 1948, pp. 481–91). The March 1954 statement of CED, which was not available at the time the article was prepared, goes beyond the *Stabilizing Budget Policy* in seeking to remove some of these limitations.

2. I am sure that it is difficult to judge the extent to which government officials are "dominated" by views about the balanced budget and by fear of the debt. Mr. Stein apparently finds hope in recent policies and statements. I have been distressed by some pronouncements emanating from the Treasury about "mortgaging our children's future" (*Life*, March 15, 1954), and would have hoped that Ricardo had dispelled some of these notions 130 years ago.

I had not thought my article particularly subtle. The "basic thesis" of the paper was an attempt to explain, in a world of Keynesianism, the prevalence of the crudest kind of Smithian views about debts and deficits. It is not only the "practical nature of the problem" but the views of those in the "sanctuary" which are cause for concern.

3. The CED has made some important contributions to public understanding of the significance of the government budget. Their popularization of the cash consolidated budget, as against the administrative budget, has been a major step forward. But I fear that Mr. Stein, and perhaps the CED, are still looking for some government budget to balance. This is an understandable but probably fruitless search. The guidelines for decisions about the transfer of resources from the private to the public sector need not and probably cannot be framed in terms of balancing a set of government accounts. I hope and trust that Mr. Stein would agree with the cliché that it is the economy, not the government accounts, which ought to be balanced.

<div align="right">Jesse Burkhead.</div>

Syracuse University

RECENT PUBLICATIONS

BAKER, HELEN AND FRANCE, ROBERT R. Centralization and Decentralization in Industrial Relations. Princeton, New Jersey; Industrial Relations Section, Princeton University. 1954. pp. 218.

BERNSTEIN, IRVING. Arbitration of Wages. Berkeley, California: University of California Press. 1954. pp. 125. $3.50.

BORENSTEIN, ISRAEL. Capital and Output Trends in Mining Industries, 1870–1948. New York, N. Y.: National Bureau of Economic Research. 1954. pp. 81. $1.00. [Occasional Paper 45.]

BOURIEZ-GREGG, FRANÇOISE. Les Classes Sociales aux Etats-Unis. Paris: Librairie Armand Colin. 1954. pp. 234.

BROIDE, JULIO. Bibliografia del Ano 1952 Sobre Ciclos Economicos. Buenos Aires, Argentina: Facultad de Ciencias Economicas, Universidad de Buenos Aires. 1953. pp. 60.

CAPRARA, UGO. La Banca: Principii di Economia delle Aziende di Credito. Milan, Italy: Antonino Giuffre. 1954. pp. 453.

CARROLL, CHARLES R. Private Investments Abroad. Washington, D. C.: American Enterprise Association. 1954. pp. 83. $1.00. [No. 454 in the series "National Economic Problems."]

CHALMERS, W. ELLISON, ET AL. Labor-Management Relations in Illini City: The Case Studies. Champaign, Illinois: Institute of Labor and Industrial Relations, University of Illinois. 1954. pp. 809. [Volume 1.]

——. Labor-Management Relations in Illini City: Explorations in Comparative Analysis. Champaign, Illinois: Institute of Labor and Industrial Relations, University of Illinois. 1954. pp. 662. [Volume 2.]

CHARDONNET, J. La Siderurgie Francaise. Paris: Librairie Armand Colin. 1954. pp. 238.

CLIFFORD, HENRY H. Investment Review and Outlook. Los Angeles, California: A. M. Clifford Associates. 1954. pp. 24.

COURT, W. H. B. A Concise Economic History of Britain from 1750 to Recent Times. New York: Cambridge University Press. 1954. pp. 368. $4.00.

DAVISSON, CHARLES N. The Marketing of Automotive Parts. Ann Arbor, Michigan: Bureau of Business Research, University of Michigan. 1954. pp. 958. [Michigan Business Studies, Volume XII, No. 1.]

EIRIKSSON, BENJAMIN H. J. Outline of an Economic Theory. Reykjavik, Iceland: Helgafell Publishing Co. 1954. pp. 493.

EKONOMSKI INSTITUT FNRJ. Tromeseeni Pregled Privrede Fnrj. Belgrade, Yugoslavia: Ekonomski Institut Fnrj. 1954. pp. 60. [Volume I, No. 1.]

FURNISS, EDGAR S., JR. The Office of the Premier in French Foreign Policy-making: An Application of Decision-making Analysis. Princeton, New Jersey: Organizational Behavior Section, Princeton University. 1954. pp. 67. [Foreign Policy Analysis Project, Foreign Policy Analysis Series No. 5.]

GLOVER, J. D. The Attack on Big Business. Boston, Massachusetts: Division of Research, Graduate School of Business Administration, Harvard University. 1954. pp. 375. $4.00.

GORIS, HENDRIEKE. Price-Determining Factors in American Tobacco Markets. Amsterdam, The Netherlands: North-Holland Publishing Company. 1954. pp. 109.

Hammond, R. J. Food and Agriculture in Britain, 1939–45: Aspects of Wartime Control. Stanford, California: Food Research Institute, Stanford University. 1954. pp. 246. $5.00.

Heurlin, Lauri O. The Economic Theory of Agricultural Production. Helsinki, Finland: The Finnish Academy of Science and Letters. 1954. pp. 130.

Hodgman, Donald R. Soviet Industrial Production, 1928–1951. Cambridge, Massachusetts: Harvard University Press. 1954. pp. 241. $5.00. [Russian Research Center Studies 15.]

Hugon, Paul. De la Méthode en Economie Politique. Sao Paulo, Brazil: University of Sao Paulo. 1954. pp. 136. [Bulletin No. 177.]

Industrial Relations Section, Princeton University. Selected References. Princeton, New Jersey: Institute of Industrial Relations. 1954. pp. 4. 20 cents. [No. 60.]

International Political Science Association. International Bibliography of Political Science. New York: Columbia University Press. 1954. pp. 248. $3.00.

Iversen, Carl. A Report on Monetary Policy in Iraq. Copenhagen, Denmark: Ejnar Munksgaard. 1954. pp. 331.

Jaffe, A. J. and Carleton, R. O. Occupational Mobility in the United States, 1930–1960. New York: Columbia University Press. 1954. pp. 105. $2.75.

Jefferys, J. B. Retail Trading in Britain, 1850–1950. New York: Cambridge University Press. 1954. pp. 497. $9.50. [National Institute of Economic and Social Research, Economic and Social Studies. XIII.]

Koyck, L. M. Distributed Lags and Investment Analysis. Amsterdam, The Netherlands: North-Holland Publishing Company. 1954. pp. 111.

Kuznets, Simon and Rubin, Ernest. Immigration and the Foreign Born. New York: National Bureau of Economic Research, Inc. 1954. pp. 107. $1.50. [Occasional Paper 46.]

Lester, Richard A. Hiring Practices and Labor Competition. Princeton, New Jersey: Industrial Relations Section, Princeton University. 1954. pp. 108. $2.50.

Lockwood, William W. The Economic Development of Japan: Growth and Structural Change, 1868–1938. Princeton, New Jersey: Princeton University Press. 1954. pp. 603. $10.00.

Lösch, August. The Economics of Location. New Haven, Connecticut: Yale University Press. 1954. pp. 520. $7.50.

Malm, F. Theodore. Recruiting Patterns and the Functioning of Labor Markets. Berkeley, California: Institute of Industrial Relations, University of California. 1954. pp. 18. [Reprint No. 60.]

Nef, John U. La Naissance de la Civilisation Industrielle et le Monde Contemporain. Paris: Librairie Armand Colin. 1954. pp. 249.

Nelson, Lowry. American Farm Life. Cambridge, Massachusetts: Harvard University Press. 1954. pp. 192. $3.75.

Papi, Giuseppe Ugo. Teoria e Politica dello Sviluppo Economico. Milan, Italy: Dott. Antonino Giuffrè. 1954. pp. 567.

Parnes, Herbert S. Research on Labor Mobility: An Appraisal of Research Findings in the United States. New York: Social Science Research Council. 1954. pp. 205. $1.75. [Bulletin 65.]

Petino, Antonio. Aspetti e Momenti de Politica Granaria a Catania ed in Sicilia nel Quattrocento. Catania, Italy: Universita di Catania. 1951. pp. 83. [Studi de Economia e Statistica, Series I, Volume II.]

——. Lo Zafferano nell'Economia del Medioevo. Catania, Italy. Universita di Catania. 1950. pp. 102. [Studi de Economia e Statistica, Series I, Volume I.]

PHELPS, CLYDE WILLIAM. Financing the Instalment Purchases of the American Family. Baltimore, Maryland: Commercial Credit Company. 1954. pp. 105. [Studies in Consumer Credit, No. 3.]

RAMAMURTI, B. Agricultural Labour: How They Work and Live. New Delhi, India: Manager Government of India Press. 1954. pp. 137.

ROBBINS, LIONEL. The Economist in the Twentieth Century. New York: St. Martin's Press, Inc. 1954. pp. 224. $3.50.

ROBSON, W. A. The University Teaching of Social Sciences: Political Science. New York: Columbia University Press. 1954. pp. 249. $1.50.

SCHILLER, OTTO. Die Landwirtschaft der Sowjetunion, 1917–1953. Tübingen, Germany: Arbeitsgemeinschaft für Osteuropaforschung. 1954. pp. 108.

SHATTUCK, LEROY A., JR. AND BRADLEY, JOSEPH F. Problems in Business Finance. New York: Rinehart and Company, Inc. 1954. pp. 67. 95 cents.

STAUB, HANS. Le Profit des Grandes Entreprises Americaines. Paris: Librairie Armand Colin. 1954. pp. 151.

TOSTLEBE, ALVIN S. The Growth of Physical Capital in Agriculture, 1870–1950. New York: National Bureau of Economic Research, Inc. 1954. pp. 92. $1.25. [Occasional Paper 44.]

UNITED NATIONS. The Economic Growth of Twenty Republics: The Work of the Economic Commission for Latin America. New York: Columbia University Press. 1954. pp. 32. 15 cents.

——. Reports and Papers on Mass Communication: Newsprint Trends, 1928–1951. New York: Columbia University Press. 1954. pp. 63. 40 cents. [No. 10.]

——. Reports and Papers on Mass Communication: Paper for Printing (other than newsprint) and Writing, 1929–1951 Trends. New York: Columbia University Press. 1954. pp. 42. 40 cents. [No. 11.]

——. Reports and Papers on Mass Communication: Paper for Printing and Writing Tentative Forecasts of Demand in 1955, 1960 and 1965. New York: Columbia University Press. 1954. pp. 105. 40 cents. [No. 12.]

——. The University Teaching of Social Sciences: Economics. New York: Columbia University Press. 1954. pp. 300. $3.00.

U. S. DEPARTMENT OF COMMERCE, BUREAU OF THE CENSUS. Statistical Abstract of the United States, 1954. Washington, D. C.: U. S. Government Printing Office. 1954. pp. 1056. $3.75.

VERNON, RAYMOND. America's Foreign Trade Policy and the GATT. Princeton, New Jersey: International Finance Section, Princeton University. 1954. pp. 25. [Essays in International Finance, No. 21.]

WICKSELL, KNUT. Value, Capital and Rent. New York: Rinehart and Company, Inc. 1954. pp. 180. $3.50.

WOODS, C. A. The Indian Community of Natal: Their Economic Position. New York: Oxford University Press. 1954. pp. 102. $4.00. [Natal Regional Survey, Volume 9.]

THE
QUARTERLY
JOURNAL OF
ECONOMICS

FOUNDED 1886

The Quarterly Journal of Economics

Editor

EDWARD H. CHAMBERLIN

Associate Editors

CONTENTS FOR FEBRUARY, 1955

SCHOLASTIC ECONOMICS:
SURVIVAL AND LASTING INFLUENCE
FROM THE SIXTEENTH CENTURY TO ADAM SMITH

By Raymond de Roover

I. Introduction: The Medieval Contribution

Shortly before the end of the nineteenth century, Luigi Cossa deplored the fact that there existed no work on scholastic economics "without some underlying bias towards systematic refutation or extravagant apology."[1] Despite Cossa's own efforts to amend this situation, there has been very little improvement in recent years, and scholastic economics has remained a field which is so neglected or so poorly cultivated that, in the opinion of most economists, it hardly deserves serious consideration. As a result, most of the standard textbooks on the history of economic thought — if they do not omit the subject altogether and start with the physiocrats — devote little space to what they call "medieval" economics. After some trite comments on Thomas Aquinas, they greet Oresme (c. 1330–1382) from a distance and then hasten on to Thomas Mun and the theory of the balance of trade. Usually, the treatment is not only superficial but replete with errors which could have been avoided by going to the sources instead of repeating clichés.[2]

1. Luigi Cossa, *An Introduction to the Study of Political Economy* (London, 1893), p. 141. Although this book is not analytical, it is still extremely useful for its bibliographical and other accurate information.

2. A laudable exception is the book of Edgar Salin, *Geschichte der Volkswirtschaftslehre* (4th ed.; Berne, 1951). Another is, of course, the great work of Joseph A. Schumpeter, *History of Economic Analysis* (New York, 1954). As this article was written independent of Schumpeter — in fact, the manuscript was sent to this *Journal* before his book appeared — no references to his *History* have been

As has already been pointed out in this *Journal,* the current textbooks entirely overlook the fact "that Aquinas was the founder of a school and that his doctrines were further elaborated and refined by his followers."[3] It should be added that these followers continued far beyond the Middle Ages until well into the seventeenth century. Moreover, some of their important economic doctrines were taken over, with only slight modifications, by the philosophers of natural law, such as Hugo Grotius (1583–1645) and Samuel Pufendorf (1622–94), who were still Aristotelians, even if they were opposed to scholasticism.

Since the later scholastic writers built on the foundations laid by their predecessors, it appears necessary to say a few words about the method used by the medieval Schoolmen and about their economic contributions of a technical nature. The author assumes that their contributions in a broader sense are known, in spite of the limited treatment accorded the subject in most histories of economic thought.

No more than the authors of antiquity, did the medieval Schoolmen consider political economy as an independent discipline, but as an appendix to ethics and law.[4] This situation still persisted in the eighteenth century when Adam Smith took charge of the chair of Moral Philosophy at Glasgow College. The courses of his predecessor, Francis Hutcheson (1694–1746), and his contemporary at Edinburgh, Adam Ferguson (1723–1816), are available in print. According to these sources, the contents of a course in Moral Philosophy in the eighteenth century and in Presbyterian Scotland still corresponded, by and large, to the description of the subject matter given in the thirteenth century by Thomas Aquinas in his *Comments on the Ethics of Aristotle.*[5] Economics, in the modern sense, occupied a very subordinate position and was still viewed as an ethical and legal matter involving the application of natural law to civil contracts.

What the Doctors in the Middle Ages were really interested in was to determine the rules of justice governing social relations. Following Aquinas, they distinguished two kinds of justice: distribu-

added. The reader may be interested in comparing this essay with Schumpeter's remarks and conclusions. He will find a different treatment of the subject, but fundamental agreement on various points.

3. R. de Roover, "Monopoly Theory prior to Adam Smith: a Revision," this *Journal* (Nov. 1951), p. 493.

4. I avoid using the term "economics" here, because in the Middle Ages it still retained the same meaning as in antiquity and referred to household management.

5. Thomas Aquinas, *In X libros ethicorum ad Nicomachum,* I, 1.

tive justice, which regulated the distribution of wealth and income, according to the place of the individual in society, and commutative justice, which applied to the reciprocal dealings between individuals, that is, to the exchange of goods and services.[6] In other words, economic matters pertained to justice, not to charity, as can be readily ascertained by merely running through the table of contents of Aquinas' *Summa theologica*.

In dealing with questions of justice, the Doctors unavoidably hit upon economic matters and were forced to consider them. At first their investigation was limited to the just price and usury, but it soon branched out to involve a host of other questions, including the just wage, debasement (inflation), justice in taxation, public debts, monopoly, foreign exchange, partnerships, and all the contracts that might involve any taint of usury.

The medieval mind was legalistic and, under the influence of Roman law, a great deal of importance was attached to the form of contracts. The principal problem was always to determine whether a contract was licit or illicit. This emphasis tended to narrow the scope of economics to the study of the legal nature of contracts and their ethical implications, a tendency which reflects itself even in the title and the arrangement of scholastic treatises. One will be sure to find economic matters discussed — along with other topics, of course — in any treatise on moral theology bearing as title *De contractibus* (Concerning Contracts) or *De justitia et jure* (Concerning Justice and Law). Almost invariably economic subjects are also touched upon in guides for confessors, though the exposition in works of this type is likely to be less systematic and analytical and more casuistic. As a matter of fact, the word "casuistry" derives from the concern of the late scholastic writers with cases of conscience.

Thomas Aquinas (1226–1274) had given a place to economics in his universal scheme: it was ruled by justice and grounded on private property and exchange. In any case, the pursuit of material welfare was not to be regarded as an end in itself, but as a means to achieve the *summum bonum* of salvation.[7] These fundamental principles were never questioned by his followers, but practical necessities soon spurred them to elaborate his rather sketchy analysis on usury and price. The first who refined it considerably was John Buridan (1300–1358), a pupil of William of Ockham and a rector of the Univer-

---

6. *Idem, Summa theologica*, II, ii, quaest. 61, arts. 1 and 2.
7. *Ibid.*, II, ii, qu. 55, art. 6, and *Summa contra Gentiles*, III, c. 30. Cf. Aristotle, *Nicomachean Ethics*, I, 5 and 8.

sity of Paris. He insisted on the point that value was measured by human wants: not by those of a single individual, but by those of the entire community (*rei venalis mensura est communis indigentia humana*).[8] He made it clear, also, that he considered the market price as the just price. Buridan's analysis even anticipates the modern concept of a consumer scale of preferences, since he states that the person who exchanges a horse for money would not have done so, if he had not preferred money to a horse.

After Buridan, the next writer of importance was the Florentine jurist and diplomat, Messer Lorenzo di Antonio Ridolfi (1360–1442), who in 1403 wrote a treatise on usury.[9] It contains the first detailed discussion of foreign exchange. Of course, he deals with the subject from the scholastic point of view, which is radically different from the later mercantilist or balance-of-trade approach.[1] The question raised by Ridolfi is whether exchange dealings are lawful or involve usury.

Lorenzo Ridolfi was followed by the famous preacher, San Bernardino of Siena (1380–1444), whom Professor Edgar Salin considers as one of the most notable economists of all times.[2] As sources of value, he lists three factors: utility (*virtuositas*), scarcity (*raritas*), and pleasurableness (*complacibilitas*). He also mentions that goods may be more or less gratifying, according to the intensity of our desire to possess and to use them. Without stretching these statements too far, it seems to me that San Bernardino had undoubtedly a psychological theory of value and even some inkling of varying degrees of utility. According to him, the just price is determined by "the estimation made in common by all the citizens of a community" (*æstimatio a communitatibus civilibus facta communiter*). In my opinion, this is clearly the competitive price in a free market. The correctness of this interpretation is beyond question, since Bernardino

8. Edmund Schreiber, *Die volkswirtschaftlichen Anschauungen der Scholastik seit Thomas von Aquin* (Jena, 1913), pp. 178–86.

9. *Tractatus de usuris et materiae montis* (1st ed.; Pavia, 1490); republished in Vol. VII of the *Tractatus universi juris* (Venice, 1583), fols. 15r–50r.

1. R. de Roover, *Gresham on Foreign Exchange; an Essay on Early English Mercantilism* (Cambridge, Mass.: Harvard University Press, 1949), pp. 173–80, and *L'évolution de la lettre de change, XIVe–XVIIIe siècles* (Paris: Armand Colin, 1953), pp. 51, 58–60, 127–29.

2. *Op. cit.*, p. 45. There are two recent monographs on the economics of San Bernardino: Franz Josef Hünermann, *Die wirtschafts-ethischen Predigten des hl. Bernardin von Siena* (Münster, 1939) and Alberto E. Trugenberger, *San Bernardino da Siena, Considerazioni sullo sviluppo dell'etica economica cristiana nel primo Rinascimento* (Berne, 1951). The sermons of Bernardino of Siena dealing with economics are in his collection, *De Evangelio Aeterno*, Nos. 32 to 42.

is outspoken in his condemnation of monopolistic practices, that is, of "fraudulent and pernicious agreements" by which merchants drive up prices in order to increase their profits. Finally, San Bernardino states that the "difficulty" of producing a good makes it scarcer and more valuable. Does he imply that the cost of production determines price by affecting the supply? An interesting point is that "difficulty," instead of supply, appears in the lectures of Francis Hutcheson, Adam Smith's teacher, as a price-determining factor. The concept is not used in *The Wealth of Nations*, but it reappears in Ricardo's *Principles of Economics* (chap. 20) where it is said that value depends upon the difficulty or facility of production, which is apparently synonymous with more or less labor. In his *Logic of Political Economy*, Thomas De Quincey (1785–1859), trying to improve upon Ricardo, recognizes two sources of value: utility and difficulty of attainment. His discussion is quoted at length and with approval by John Stuart Mill in his *Principles of Political Economy* (Book III, chap. 2, §1). These observations lead to two conclusions. First, the persistent use of the same terminology points to a continued tradition. Second, it seems that this part of value analysis made little, if any, progress from the time of San Bernardino to John Stuart Mill. On the contrary, it might even be argued that the latter's analysis is inferior, because it is less explicit on the point that difficulty creates scarcity.

Although San Bernardino, like the other Schoolmen, regards money as sterile, he contradicts himself elsewhere when he admits that it acquires a seminal quality by becoming "capital."[3] By capital, he does not mean the principal of a debt, but money invested in a business venture.[4] The same contradiction is found in Thomas Aquinas, who, in one passage, affirms that money is barren and, in another, compares it to seed which, if put into the soil, will sprout and produce a crop.[5]

San Bernardino also mentions *cambium* and government debts, but a better discussion of these topics is found in the writings of his contemporary, San Antonino (1389–1459), Archbishop of Florence.[6]

3. Ernest Nys, "The Economic Theories of the Middle Ages," *Researches in the History of Economics* (London, 1899), p. 164.

4. In this meaning, the word "capital" is already used in notarial and business records from the twelfth century onward. Numerous examples occur in the cartulary of Giovanni Scriba, or John the Scribe, which contains acts dating from 1154 to 1164 (*Il cartolare di Giovanni Scriba*, ed. Mario Chiaudano, Turin, 1935).

5. *Summa theologica*, II, ii, qu. 61, art. 3. Elsewhere (qu. 78, art. 1), quoting Aristotle, Aquinas states that money was invented mainly to serve as a medium of exchange.

6. On San Bernardino and San Antonino, see also Amintore Fanfani, *Le origini dello spirito capitalistico in Italia* (Milan, 1933), pp. 106–19.

Although not a very original thinker, San Antonino wrote with ease and was well versed in the extant canonistic and theological literature. His works contain an excellent summary of the controversy, then raging, about the lawfulness of interest-bearing shares in the public debt. With regard to value and price, he takes over the theory of San Bernardino without modification; yet he has often received undeserved credit as the first to mention utility.[7]

The last of the important economic writers of the Middle Ages is Thomas de Vio (1468–1524), better known as Cardinal Cajetan.[8] His work in the field of economics includes three brief treatises: one on usury, another on *cambium*, and a third on the *Montes Pietatis*, which he bitterly opposed. The most remarkable of the three treatises is perhaps the one on *cambium*, in which he shows himself well-informed on banking practices. In conformity with scholastic dialectics, he defines *cambium* as a *permutatio*, one of the nominate contracts found in Roman law, and not as a *mutuum*. Thus he justified real exchange provided that the place difference be observed, that is, that the bill of exchange be issued in one place and payable in another. Dry exchange, a practice without analogy in modern business, is proscribed because it is a faked exchange transaction violating this rule.[9]

This brief and incomplete survey omits the minor Schoolmen, some of whom are not without interest. The authors discussed are all men of singular merit, justly famous for their achievements, not merely in economics but chiefly outside this field.

A grave shortcoming of the medieval as well as the later Schoolmen is their overemphasis of the usury question. The space devoted to it in scholastic treatises has given the mistaken impression that it was regarded as all important. Sir William Ashley even asserts that "the prohibition of usury was clearly the centre of the canonist doctrine."[1] This is untrue. As stated above, the Schoolmen considered equity in distribution and exchange as the central problem in eco-

---

7. On the economic doctrines of San Antonino, there are the following studies, none of outstanding quality: Carl Ilgner, *Die volkswirtschaftlichen Anschauungen Antonins von Florenz* (Paderborn, 1904); Bede Jarrett, *San Antonino and Mediaeval Economics* (St. Louis, 1914); and August Pfister, *Die Wirtschaftsethik Antonin's von Florenz (1389–1459)* (Fribourg, Switzerland, 1946).

8. *De Monte Pietatis* (1498), *De cambiis* (1499), *De usura* (1500); republished recently in *Scripta philosophica, opuscula oeconomico-socialia*, ed. P. P. Zammit (Rome, 1934).

9. R. de Roover, "What is Dry Exchange? A Contribution to the Study of English Mercantilism," *Journal of Political Economy*, LII (1944), 250–66.

1. *An Introduction to English Economic History and Theory*, Vol. I, Part 2 (9th impression, London, 1920), p. 395. Cf. *ibid.*, p. 382.

nomics. The usury question was a side issue, but concern with it was allowed to crowd out nearly everything else.[2]

II. The School of Salamanca

By many authors, Gabriel Biel (*c.* 1435–1495), professor at the University of Tübingen, is considered the last of the Schoolmen. Actually, scholasticism did not die with him; on the contrary, it received a new lease on life in the sixteenth century. This regeneration was the work of the school founded by Francisco de Vitoria (1480–1546), who, from 1526 to 1544, taught at Salamanca — in this period, the queen of the Spanish universities. As a matter of fact, the term "the school of Salamanca" is often applied to the body of his students, his disciples and their successors.[3] From Spain, the influence of Francisco de Vitoria's teaching spread to Portugal (to the University of Coïmbra), to Italy (through the Roman college of the Jesuits), and to the Low Countries where Leonardus Lessius (1551–1623), Franciscus Sylvius or du Bois (1581–1649), and Johannes Malderus (1563–1633) wrote commentaries on Thomas Aquinas inspired by the Spanish doctrines.

The school of Salamanca distinguished itself in philosophy and in natural and international law. The treatises of Francisco de Vitoria on the Indies and on the laws of war have even been republished by the Carnegie Endowment for International Peace.[4] Some of Vitoria's pupils occupied prominent positions: Domingo de Soto (1494–1560) represented Charles V at the Council of Trent and in 1548 became the Emperor's confessor; Diego de Covarrubias y Leyva (1512–1577), who wrote a treatise on money, was appointed Bishop of Ciudad Rodrigo and later President of the Council of Castile;[5] Martín de Azpilcueta, better known as Navarrus (1493–1586), was rector of the University of Coïmbra before being called to Rome in 1567, where he enjoyed the confidence of three successive popes and died a nonagenarian.[6] Among those influenced indirectly by Francisco

2. On usury, by far the best study available in English is the article of T. P. McLaughlin, "The Teachings of the Canonists on Usury (XII, XIII and XIV Centuries)," *Mediaeval Studies*, I (1939), 81–147, and II (1940), 1–22. Cf. Benjamin N. Nelson, *The Idea of Usury* (Princeton: Princeton University Press, 1949).

3. Marjorie Grice-Hutchinson, *The School of Salamanca, Readings in Spanish Monetary History, 1544–1605* (Oxford: The Clarendon Press, 1952). Although not written by a professional economist, this is an excellent little book.

4. *De Indis et de jure belli: Relectiones* (Washington, D. C., 1917). No. 7 of the series: The Classics of International Law.

5. *Veterum numismatum collatio* (Salamanca, 1550).

6. On Azpilcueta there is a study by Alberto Ullastres Calvo, "Martín de Azpilcueta y su comentario resolutorio de cambios; las ideas ecónomicas de un moralista español del siglo XVI," *Anales de Economía*, I (1941), 375–407, and II (1942), 52–95.

de Vitoria, mention should be made also of Luis de Molina (1535–1601), who occupied for more than twenty years the chair of theology at the University of Evora in Portugal. His analysis of value and price is especially valuable for its comprehensiveness.[7] Since economics was not recognized as an independent discipline, it is not surprising that the members of the school of Salamanca achieved greater distinction in other fields, but this is no reason why they should be ignored by economists or denied their due.

In form and content, the treatises published by the Spanish school continue the scholastic tradition with its constant appeal to authority, its display of references to support even the most trivial statements, and its love of subtle distinctions and definitions.[8] As in the past, attention remained focused on the observance of the rules of justice and on the lawfulness of various types of contracts. The moralists of the new school, however, attempted to provide fresh interpretations, to refine their concepts, to elaborate their analysis, to observe market conditions, and to bring their principles somehow into harmony with the requirements of expanding business and finance. Without changing completely the scholastic methods of analysis, the task was by no means an easy one. No wonder that the casuists of the Spanish school were only half successful; they revitalized scholasticism, it is true, but only for a time, without saving it from ultimate doom.[9]

Even more than the medieval Schoolmen, the later Doctors adhered to the theory that utility was the main source of value and that the just price, in the absence of public regulation, was deter-

---

7. Bernard W. Dempsey, *Interest and Usury*, with an Introduction by Joseph A. Schumpeter. This work discusses the theories of three Schoolmen, all belonging to, or influenced by, the school of Salamanca: Molina, Lessius and Lugo. On Molina, there is an unpublished doctoral dissertation by W. Seavey Joyce, *The Economics of Luis de Molina* (Harvard University, 1948).

8. Grice-Hutchinson, *op. cit.*, p. 40. Nevertheless, Earl J. Hamilton (*American Treasure and the Price Revolution in Spain, 1501–1650*, p. 295) labels Tomás de Mercado and others as "Spanish mercantilists." This label is certainly wrong: even the title of Mercado's treatise, *Summa de tratos y contratos de mercaderes* (1st ed.; Salamanca, 1569), indicates clearly that the approach is scholastic. Moreover, Spanish writers do not consider Mercado and other authors belonging to the school of Salamanca as mercantilists, but call them *jusnaturalistas* and *moralistas*, which is correct, in my opinion: José Larraz, *La época del mercantilismo en Castilla, 1500–1700* (2d ed.; Madrid, 1943), pp. 119, 122, and 131. Cf Andrés V. Castillo, *Spanish Mercantilism: Gerónimo de Uztáriz, Economist* (Columbia University Press, 1930), p. 45.

9. With reference to the philosophical restoration initiated by the school of Salamanca, the same views are expressed by Maurice De Wulf, *History of Mediaeval Philosophy*, II, 301–7.

mined by common estimation, that is, by the interplay of the forces of supply and demand without any frauds, restraints, or conspiracies.[1] Domingo de Soto and Luis de Molina both denounce as "fallacious" the rule formulated by John Duns Scotus (1274–1308), according to which the just price should equal the cost of production plus a reasonable profit.[2]

Tomás de Mercado makes the pertinent remark that prices are as changeable as the wind.[3] Molina even introduces the concept of competition by stating that "concurrence" or rivalry among buyers will enhance prices, but that a flagging demand will bring them down.[4] Since similar statements are found in other authors, we may conclude that the Doctors of the new school generally accepted the idea that the just price, if not fixed by public authority, corresponded to the current or market price.[5]

Conditions of supply and demand are not the only factors that affect prices. There is also the influence of the volume of circulating media on the price level. The Spanish authors take the quantity theory for granted, since their treatises, almost without exception, mention that prices go up or down according to the abundance or scarcity of money.[6] Twelve years before Jean Bodin, or in 1556, Azpilcueta or Navarrus, attributes the rise of Spanish prices to the

1. Luis de Alcalá, *Tractado de los prestamos que passan entre mercaderes y tractantes* (Toledo, 1546), Part I, § 5, fol. 5ᵛ; Part II, § 11, fol. 22–23; Luis Saravía de la Calle, *Instrución de los tratos del comprar y vender* (Medina del Campo, 1544), cap. 2; Tomás de Mercado, *op. cit.*, lib. 2, cap. 8; Domingo de Soto, *De justitia et jure* (1st ed.; Salamanca, 1553), lib. VI, quest. 2, art. 3; Luis de Molina, *De justitia et jure* (Cuença, 1592), tract. II, disp. 348, § 8. Cf. Grice-Hutchinson, *op. cit.*, pp. 49, 72, 79, 82, 88. Soto expresses the rule as follows: *Sensus ergo est quod tantum valet res quanti vendi potest, seclusa vi, fraude et dolo.*

2. Soto, *op. cit.*, lib. VI, qu. 2, art. 3; Molina, *op. cit.*, tract. II, disp. 348, §8; and Mercado, *op. cit.*, lib. 2, cap. 11. Cf. Bernard W. Dempsey, "Just Price in a Functional Economy," *American Economic Review*, XXV (1935), 471–86.

3. *Op. cit.*, lib. 2, cap. 8: . . . *Aunque es mas variable (según la experiencia enseña) que el viento.*

4. *Op. cit.*, Tract. II (*de contractibus*), disp. 348, §4: *Multitudo emptorum concurrentium, plus uno tempore quam alio, et maiore aviditate facit pretium accrescere; emptorum vero raritas facit illud decrescere.*

5. Grice-Hutchinson, *op. cit.*, pp. 48, 80, 86–87, 105.

6. Molina, for example, states that prices and wages will be higher in a country where money is abundant than in another where it is scarce (*op. cit.*, tract. II, disp. 348, §4). Cf. Grice-Hutchinson, *op. cit.*, pp. 80, 105; Mercado, *op. cit.*, lib. 2, cap. 11; Cardinal de Lugo quoted by J. Brodrick, *The Economic Morals of the Jesuits*, p. 10. Cf. Bernard W. Dempsey, "The Historical Emergence of Quantity Theory," this *Journal*, L (1936), 174–84, and the "Comments" thereto by E. J. Hamilton, *ibid.*, 185–92. These comments only illustrate the fact that economists look in the wrong places for bibliographical guidance on scholastic economics.

influx of gold and silver from the New World.[7] He also observes that, because the flow reaches Spain first, the level of prices and of wages is higher there than in France.

The Spanish moralists devoted much more attention to foreign exchange than did the medieval Schoolmen. They noticed that in the trade with Flanders and Italy, the exchange rate was generally unfavorable to Spain, but they could not explain this phenomenon, since they ignored the balance-of-payments theory.[8] Instead, they sought to justify exchange transactions by arguing that the money had a greater purchasing power abroad than in Spain and that the rate had to be unfavorable in order to restore the purchasing power parity, a partial and ambiguous explanation, but typical of scholastic dialectics.[9] Among the Spanish moralists, a lively discussion was raised concerning the lawfulness of exchange between two places within the same realm. This practice, it was charged, served merely as a subterfuge to circumvent the usury prohibition.[1] In the debate the rigorist friars, headed by Domingo de Soto, eventually won out, and, through their influence at Court, secured in 1552 a royal decree forbidding internal exchange at any rate other than par. It is needless to say that the merchants soon discovered new ways of evasion. In trying to constrain the market, the moralists were fighting a losing battle.

One of the major developments of the sixteenth century was the rise of the fairs of Castile, Lyons, Frankfort-on-the-Main, and, above all, Besançon, as international clearing centers. From 1579 on, the Besançon fairs, while keeping their name, were actually held in

---

7. *Comentario resolutorio de cambios* (Salamanca, 1556), cap. 20, no. 51. For an English version of this passage, see Grice-Hutchinson, *op. cit.*, p. 95. Molina, whose text is, however, posterior to the book of Bodin, also refers to the price-raising effects of the influx of gold and silver from New Spain (*op. cit.*, tract. II, disp. 83, §13). Azpilcueta, or Navarrus, is not mentioned in the works of Hamilton.

8. Mercado, *op. cit.*, lib. IV, cap. 4; Soto, *op. cit.*, lib. VI, qu. 12, art. 2. Cf. de Roover, *L'évolution*, p. 81; Grice-Hutchinson, *op. cit.*, pp. 13–14.

9. *Ibid.*, pp. 57–58. Of course, this is not the purchasing-power-parity doctrine as understood today.

1. Those opposed to internal exchange were Francisco de Vitoria and Domingo de Soto (*op. cit.*, lib. VI, qu. 13, art. 1). On the other hand, Miguel de Palacios and Tomás de Mercado (*op. cit.*, tract. 4, cap. 8) considered it lawful. Francisco Garcia (*Trattato di tutti i contratti che nei negotii et commertii humani sogliono occorrere*, Brescia, 1596, cap. 36, §7), without being pro or con, simply states that, in Spain, internal exchange is prohibited by law. The same position is taken by Azpilcueta, or Navarrus (*op. cit.*, cap. 15, nos. 28–30), who is, however, very sceptical about the practical results of the prohibition. Cf. de Roover, *L'évolution*, pp. 108, 184, 195, 200, 202, 205.

Piacenza, on the initiative of the Genoese bankers who monopolized the financial business of the Spanish crown. To a certain extent, these fairs were institutions called forth by the scholastic doctrine, since it condemned the discounting of credit instruments but did not frown upon dealings in foreign exchange, unless they were overtly misused to evade the ban against usury. Thus the exchange business at the fairs became one of the main preoccupations of the moralists. The copious works of two Italians, Sigismondo Scaccia (*c.* 1568–1618) and Raphael de Turri (*c.* 1578–1666), not to speak of minor treatises, deal exclusively with this topic. The principal bone of contention was the lawfulness of the *cambio con la ricorsa*, a device which involved drafts and redrafts going back and forth between Genoa or another banking place and the fairs of Besançon.[2] To befuddle the theologians, the bankers had shrouded the *cambio con la ricorsa* in a veil of technical jargon and complicated bookkeeping. Stripped of its trappings, the *cambio con la ricorsa* loses all its mystery: in its naked form it is simply discount cleverly concealed under the form of fictitious exchange transactions. Nevertheless, the theologians and the jurists, approaching the problem from a legal point of view, found themselves caught in a web of technicalities and contradictions which contributed not a little to the discredit of scholastic economics.

In economics, the scholastic doctrine reaches its full maturity in the monumental works of Cardinals Juan de Lugo (1583–1660) and Giambattista de Luca (1613–1683), who should not be mistaken one for the other, although the similarity in name leads to confusion.[3] Despite an impressive display of scholarship, their works ill conceal the fact that the Doctors had exhausted the possibilities of their method and that further progress no longer depended upon more elaboration and refinement, but upon a complete renewal of the analytical apparatus.

III. The Demise of Scholastic Economics

The demise of scholasticism is not limited to economics, of course, but involves the entire scientific and philosophical system born in the medieval universities and still far from moribund on the eve of the seventeenth century. The scepticism of the Renaissance, it is true, had sapped the strength of the scholastic system but without being

2. *Ibid.*, pp. 80–81, for an example of *cambio con la ricorsa.* Further information is found in the recent book of Giulio Mandich, *Le pacte de Ricorsa et le marché italien des changes au XVIIe siècle* (Collection "Affaires et Gens d'affaires," No. 7, Paris: Armand Colin, 1953).

3. The work of Cardinal de Lugo, *Disputationes scholasticae et morales* (Lyons, 1642), was republished in 1869. Volume VII (*in quo de contractibus in*

able to destroy the still vigorous organism. Although derided and ridiculed by its opponents, scholasticism continued to exert far-reaching influence. It was confronted, however, with an increasingly hostile spirit, which provided a favorable climate for the reception of Cartesian philosophy. The real crisis did not come until the seventeenth century. In the face of the attack, the Aristotelians failed to realize that, in order to survive, they had to renew their methods. Instead, they stubbornly refused to accept the new discoveries in experimental science, with the inevitable result that their philosophy shared the fate of their antiquated astronomy, physics, and medicine, and along with them, fell into complete discredit.[4]

On the continent of Europe, and to a lesser extent in England, the dying Aristotelian system kept its hold on the universities, which thus became asylums for old fogies and citadels of bigoted pedantry. Learning deserted this musty environment and found a haven in the academies and in the salons of the eighteenth century.

It would be a grievous mistake to view the evolution of economics as divorced from that of the other sciences. The main reason why scholastic economics decayed was that its adherents were unable or unwilling to revamp their system and to discard the dead wood in order to preserve what was worth preserving. Nothing illustrates this failure better than the work of the late casuists of the seventeenth century, such as the treatise of Raphael de Turri. In it, the scholastic doctrine on the *cambium* contract reached maturity, but the subtle distinctions between licit and illicit exchange fail to cover up the fallacies and the inconsistencies which underlie the whole argument. Why should one form of exchange be lawful and not another? One can only agree with the mercantilist Malachy Postlethwayt, who in 1751 declared that the lawyers and divines with "their useless niceties" and "their fanciful divisions and subdivisions," instead of clearing up the matter, had "only perplexed and confounded it."[5] Already in the sixteenth century, the Dominican friar, Domingo de Soto, had sounded the alarm by stating that "the matter of exchange, although

genere et in specie agitur) contains the part dealing principally with economics. Cardinal de Luca wrote a popular work in the vernacular, *Il dottor volgare* (Rome, 1673, 9 vols.), which, as the title indicates, was designed to explain the doctrine of the Doctors to the public. He is also the author of a Latin treatise, *Theatrum veritatis et justitiae* (Rome, 1669–1681, 21 vols.) written only for scholars.

4. Maurice De Wulf, *op. cit.*, II, 309 ff. The late author, a pupil of Cardinal Mercier, was professor at the Catholic University of Louvain and at Harvard University.

5. "Bill of Exchange," *The Universal Dictionary of Trade and Commerce* (2d ed.; London, 1757), p. 277.

sufficiently abstruse in itself, is being more and more obscured by the clever subterfuges of the merchants and the contradictory opinions of the Doctors."[6] But he himself was a prisoner of his method and could not escape from the impasse.

There was nothing basically wrong with the scholastic theory on value and price. It rested on utility and scarcity, and Adam Smith did not improve upon it.[7]

The great weakness of scholastic economics was the usury doctrine. Canon law, dating as it does back to the early Middle Ages when most loans were made for consumption purposes, defined usury as any increment demanded beyond the principal of a loan. Since this definition was a part of Catholic dogma, the Schoolmen were unable to change it. As time went by, it became a source of increasing embarrassment. Tied to their definition, the Doctors were sucked deeper and deeper into a quagmire of contradictions. It is not that the Church ever seriously hampered business investments, but practical necessity placed before the moralists the well-nigh impossible task of legitimizing means for taking interest while safeguarding the principle that loans were gratuitous contracts. This difficulty was solved in two ways: (1) by the doctrine of extrinsic titles, and (2) by the rather artificial distinction between licit and illicit contracts. In the sixteenth century, the more lenient among the casuists undermined their own position still further by permitting the triple contract, according to which the borrower guaranteed to the lender a fixed return of, let us say, 5 per cent a year.[8] In the end, the lawfulness of interest became a question of formality, that is, of drafting contracts in the proper form. Is it then surprising that casuistry acquired such a bad connotation and is today synonymous with sophistry and mental reservation?

6. *Op. cit.*, lib. 6, qu. 8, art. 1. This text was copied by other Doctors, see de Roover, *L'évolution*, p. 72.

7. Emil Kauder, "Genesis of the Marginal Utility Theory: From Aristotle to the End of the Eighteenth Century," *Economic Journal*. LXIII (1953), 638–50; *idem*, "The Retarded Acceptance of the Marginal Utility Theory," this *Journal*, LXVII (1953), 564–75.

8. The triple contract, as the name indicates, involves a combination of three contracts in one: (1) a partnership contract between the lender and the borrower sharing in profit and loss of the borrower's business, (2) an insurance contract by which the borrower guarantees restitution of the capital, and (3) another insurance contract by which the borrower guarantees the lender against any loss, if the latter foregoes his share in eventual profits, in exchange for a fixed but reduced return on his investment. Although the triple contract had been condemned, in 1586, by Sixtus V (1585–1590), the casuists continued to debate its lawfulness throughout the seventeenth century. "Usure," *Dictionnaire de Théologie Catholique*, XV (1948), cols. 2373–74.

From the start, the usury doctrine became an easy target for the opponents of scholasticism. In a certain way, the Doctors had only themselves to blame: by their inconsistencies, they had exposed themselves to criticism and even ridicule.

The firing was opened in the sixteenth century with the fierce attack of Charles du Moulin (1500–1566), who advocated the toleration of a moderate rate of interest. He pointed out that the usury prohibition, meant to protect the debtor, had the opposite effect by increasing the cost of lending. Not content with marshaling serious arguments, he poked fun at the "jargon" of the Doctors and at their classification of *cambium* into real, dry, and fictitious exchange, rechange and "counter-change."[9] Du Moulin's book was premature and exposed its author to persecution for heresy.

In the seventeenth century, the blast came from another quarter; this time the casuists were criticized, not for their rigor but for their leniency, by the Jansenist, Blaise Pascal (1623–1662), illustrious philosopher, mathematician, and physicist, who, by the excellency of his style, won a name for himself in French literature. His *Lettres Provinciales* were scurrilous pamphlets, that created an enormous sensation. In the eighth letter, he attempts to confute the casuists for their opinion on usury and contracts. Of course, Pascal was an amateur in economics, as well as in theology. Nevertheless, there is no denying that his castigation was not entirely undeserved and that his victims had made concessions inconsistent with their basic principles.[1]

During the eighteenth century, the attack continues unabated. Whenever the *Philosophes* refer to the Doctors, they call them "casuists" with an undertone of scorn and contempt. They refer to them only to criticize; and when they borrow from them, they do not give them any credit. This attitude is typical of the Age of Enlightenment, which showed no appreciation of Gothic cathedrals or of things medieval, in general.

More than ever, the usury doctrine is the center of attack. According to Turgot (1727–1781), the prejudice against interest had been introduced "in centuries of ignorance" by theologians who did

9. Charles du Moulin, *Sommaire du livre analytique des contracts, usures rentes constituées, intérests et monnoyes,* in *Omnia quae extant opera* (Paris, 1681), Vol. II, No. 73: "Je laisse aussi leurs jargons et distinctions de change réal, fict et sec, rechange et contre-change." Cf. de Roover, *L'évolution,* p. 195.

1. To be specific, I refer to the comments of Pascal on the Mohatra contract, one of the subterfuges used in Spain. Pascal's main victim was the casuist, Antonio Escobar y Mendoza, whose major work is *Universae theologiae moralis receptores* (1st ed.; Lyons, 1652, 7 vols.).

not "understand the meaning of the Scriptures any better than the principles of natural law."[2] Richard Cantillon (d. 1734) remarked sarcastically that *lucrum cessans* would entitle a man making "five hundred per cent" in his business to demand the same from his borrowers.[3] Abbé de Condillac frankly comes out with the assertion that the loan at interest is just and should be permitted. He goes on to state that legislators and "casuists" are confused on the subject and asks them pointedly why they disapprove of interest and not of exchange.[4] Is there really so much difference between distance in time (*distance de temps*) and distance of place (*distance de lieu*)? In France, during the eighteenth century, the law still proscribed the taking of interest, although this practice was generally tolerated with the connivance of the courts. The physiocrats waged an unsuccessful campaign for the enactment of a statute which would legalize contractual clauses stipulating the payment of interest. Such a law was not passed until October 12, 1789, after the outbreak of the French Revolution.

In Italy, the attack on the usury doctrine was even more insidious than in France, since it was carried on under the cover of orthodoxy. In 1744, there appeared a book in which the author, Marquess Scipione Maffei (1675–1755), pretended, on the surface, to defend the traditional doctrines, but, in the last chapters, ruined the entire edifice by advocating a new extrinsic title, never admitted by the theologians: the *lex principis*, that is, the law or custom of the land.[5] As a matter of fact, the purpose of the book was to justify the issue by the city of Verona of a municipal loan yielding interest at 4 per cent. Maffei's publication created such a stir that in order to appease the tempest, Pope Benedict XIV (1740–1758) was impelled to promulgate the encyclical *Vix Pervenit* (1745), which reasserted for the last time the old dogma with respect to usury.[6] Within a few months (1746), there appeared the second edition of Maffei's book without

2. "Réflexions sur la formation et la distribution des richesses," *Oeuvres de Turgot*, ed. Gustave Schelle, II (Paris, 1914), 577–78, §LXXIII:" Erreurs des scolastiques réfutées"; and "Mémoire sur les prêts d'argent" (1770); *ibid.*, III (Paris, 1919), 163. Cf. Jean François Melon, *Essai politique sur le commerce* (Paris, 1761), pp. 259 and 272.

3. *Essai sur la nature du commerce en général*, ed. Henry Higgs (London, 1931), pp. 208–10.

4. *Le commerce et le gouvernement considérés relativement l'un à l'autre*, in *Mélanges d'économie politique*, ed. Eugène Daire (Paris, 1847), I, 311–12.

5. *Dell'impiego del danaro, libri tre* (1st. ed.; Rome, 1744; 2d ed.; Rome, 1746).

6. de Roover (*L'évolution*, pp. 123–24 n.) gives a summary in French of the five points discussed in *Vix Pervenit*.

any substantial modification of the author's stand on the usury question. Yet this second edition published in full the text of the Encyclical, bore the imprimatur of the ecclesiastical authorities, and contained a dedicatory letter to Benedict XIV, a personal friend of the author. On scholasticism, the book of Maffei had a deleterious effect, since it implicitly redefined usury as any increment — not beyond the principal — but beyond *the moderate rate allowed by law or custom*.[7]

The new definition represented a radical departure from the basic norms of scholastic economics.[8] Books challenging the thesis of Maffei and restating the scholastic tradition were still being written in the beginning of the nineteenth century, but their authors were not men of any talent and they repeated the old, worn-out arguments without contributing anything new.[9] Scholasticism had ceased to attract the best minds: its discredit, except in ultra-conservative circles, was too profound.

After the *Code Napoléon*, adopted all over western Europe, had allowed the taking of interest, the Church, too, decided to abandon the old usury doctrine. It was quietly buried in 1830, when the Sacred Penitentiary issued instructions to confessors not to disturb penitents who lent money at the legal rate of interest without any title other than the sanction of Civil Law.[1] With this decision,

7. The same thesis was defended by the Jansenist, Nicolas Broedersen, in his book, *De usuris licitis et illicitis* (1st ed.; 1743). Abbé Ferdinando Galiani in his book, *Della moneta*, devotes to usury an equivocal chapter, in which he pays at least lip service to the traditional doctrine. See Arthur Eli Monroe, *Early Economic Thought, Selections from Economic Literature prior to Adam Smith*, pp. 300–7. Antonio Genovesi (1713–1769) in his book, *Lezioni di economia civile*, adopts the same point of view as Maffei and Broedersen.

8. According to the old canon law, any statutes allowing interest were anti-canonical. The Council of Vienne (1311–1312) explicitly declares them to be in contravention of divine and human law and orders the repeal of those in operation: *Decretales*, c. *Ex gravi*, in Clement., lib. 5, title 5, can. 1, §1.

9. Count Monaldo Leopardi (1776–1847), *La giustizia nei contratti e l'usura* (Modena, 1834); Anonymous, *Analisi ragionata e critica dei libri tre su le usure dell'abbate Marco Mastrofini data in luce da un amico della verità* (Naples, 1835). Count Leopardi was the father of the famous poet, Giacomo Leopardi: his reactionary ideas were not limited to the usury question.

1. "Usurc," *Dictionnaire de Théologie Catholique*, XV, cols. 2379 f. The new canon law, promulgated by Benedict XV in 1917, art. 1543, admits the validity of the legal title (*non est per se illicitum de lucro legali pacisci, nisi constet ipsum esse immoderatum*), although it still upholds the principle that a loan is *per se* a gratuitous contract. Further details about the decisions of the Roman congregations in 1830 and later may be found in Henry C. Lea, "The Ecclesiastical Treatment of Usury," *Yale Review*, II (1893–94), 379–85 (republished in *Minor Historical Writings and Other Essays*, Philadelphia, 1942, pp. 129–51). The interpretations of this author are generally unfavorable to the Church, but his factual information agrees with the *Dictionnaire de Théologie Catholique*.

scholastic economics, which had emphasized usury so much, received its death blow.

IV. Scholasticism and Mercantilism: A Contrast

The differences between mercantilism and scholastic economics are striking and profound. Yet, I do not know that a comparison has ever been attempted, although a clear perception of the contrasts has its importance for an understanding of the development of economic thought. There are even historians who profess to find the "prehistory" of economics among the vagaries of the mercantilistic pamphleteers, thus completely ignoring the contributions of the Doctors.[2]

Unlike mercantilism, scholastic economics enjoyed the unquestioned superiority of being an integral part of a coherent philosophical system. Although economics was not yet acknowledged as an independent discipline, it formed a consistent body of doctrine according to which economic relations ought to be ruled by the laws of distributive and commutative justice. In contrast, mercantilism was never more than a conglomerate of unco-ordinated prescriptions by which the authors of the mercantilistic tracts sought to influence economic policy, usually in a sense favorable to their private interests.[3]

The Doctors, as this name indicates, were all university graduates, trained in theology or in canon and civil law (*doctor utriusque juris*). Most of them were clerics, though there are some notable exceptions among the jurists, especially among the civilians, for instance, Messer Lorenzo di Antonio Ridolfi, who was a layman, a diplomat and a lecturer at the Florentine athenaeum.[4] The mercantilists, on the contrary, were with few exceptions self-trained merchants, with some literary talents, but without university degrees. Essentially, they were empiricists who, for better or for worse, were not encumbered by scholastic traditions. In this way they made their major contribution by developing the balance-of-trade theory, whereas the Doctors were unable to cut themselves loose from their traditional approach to the foreign exchange problem.

As a rule, the mercantilist writings were brief tracts on specific and controversial issues, which contrast markedly with the weighty and often pedantic treatises of the Doctors. Whereas the mercan-

2. For instance, Edward Heimann, *History of Economic Doctrines*, pp. 22–47.
3. A. V. Judges, "The Idea of a Mercantile State," *Transactions of the Royal Historical Society*, 4th Series, XXI (1939), 50.
4. For his biography, see Vespasiano da Bisticci, *Vite di uomini illustri del secolo XV* (Florence, 1938), pp. 401–5.

tilist pamphlets rarely refer to sources or provide marginal notes, the scholastic treatises literally bristle with references in support of nearly every statement, even the most commonplace. This sometimes annoying display of erudition, first introduced by the postglossators, received further encouragement from the humanists, who developed the habit of invoking the authority of the Ancients for everything.

By the very fact that the Doctors were moralists, their main preoccupation was with social justice and general welfare, but naturally with these ideals as they were conceived in the Middle Ages and the sixteenth and seventeenth centuries. The mercantilists, too, professed to further the cause of the commonweal; however, their declarations in this respect should not always be taken at their face value. All too often they serve as a screen for private interests. Most of the authors of mercantilist tracts had an ax to grind. This is especially true of the early mercantilists. Gérard de Malynes (fl. 1586–1641) was a perennial office-seeker who advocated exchange control in the hope that he himself would be appointed the controller. Misselden (fl. 1608–1654) and John Wheeler (fl. 1601–1608) were spokesmen for the Merchant Adventurers; and Thomas Mun (1571–1641) wrote his tracts in defense of the East India Company. As for Gresham (1519–1579), he was a shrewd and none too scrupulous manipulator of the money market, whose recommendations, although advantageous to the Queen, were apt to have unfavorable repercussions on English trade and on the volume of employment. The later mercantilists were less prejudiced, but their views were still warped by their narrow nationalism. Most of them rallied to the defense of the colonial system and sponsored aggressive measures to combat or to exclude foreign competition, an attitude which is alien to the spirit of scholasticism. Did not St. Thomas justify international trade by pointing out the fact that no nation is self-sufficient?[5]

As we have seen, the casuists of the seventeenth century were either unwilling or unable to rejuvenate their methods. They continued in the old ruts and made no effort to incorporate new discoveries, such as the balance-of-trade theory, into their traditional doctrines. The conservatism of the late scholastic writers thus became an impediment to further progress, and it is fortunate that the mercantilists displayed more initiative and did not hesitate to blaze new trails. True, their methods were not always sound, nor

5. Amintore Fanfani, *Storia delle dottrine economiche: il volontarismo* (3d ed.; Milan, 1942), p. 112. The reference is to *De regimine principum*, Book 2, chap. 3.

always successful, but they opened up new avenues for further research. The controversy of the early mercantilists about exchange control led to a premature proposal for the creation of a stabilization fund and eventually culminated in the formulation by Thomas Mun of the balance-of-trade theory.[6] The mercantilists also made the first clumsy attempts to use statistical data, and Sir William Petty (1623–1687) even made statistics the basis of his *Political Arithmetick*. Others pondered over banking schemes; and the studies of Charles Davenant (1656–1714) and Gregory King (1648–1712) on the behavior of grain prices put them on the track of the elasticity of demand.[7] The seventeenth century was the age of projectors. Nearly always, the aim was to influence public policy, whereas the scholastic writers were content to set up ethical standards, but left their practical realization to the often inefficient government authorities.

The scholastic writers regarded trade as an occupation which, although not evil in itself, endangered the salvation of the soul, as the merchants almost unavoidably succumbed to the temptations of usury, cheating, and unlawful gain: *et de hoc rarissime evadunt mercatores*, as St. Bonaventure (1221–1275), the Seraphic Doctor, testifies.[8] In this opinion, the other Doctors concur: without exception, they much prefer agriculture to trade. The mercantilist writers, of course, take exactly the opposite point of view.[9] In their eyes trade is the noblest of all professions.[1] Both agriculture and industry depend on trade to provide a market for their products and to give employment to the "poor."[2] The merchant, far from being regarded

6. de Roover, *Gresham on Foreign Exchange*, pp. 226–31, 250–65.

7. I take advantage of this opportunity to call the attention of the economists to an article by Luigi Einaudi, "La paternità della legge detta di King," *Rivista di storia economica*, VIII (1943), 33–38. The author attributes to both Davenant and King the discovery of the law stating that grain prices vary more than proportionately to the deviations of the harvest from the normal.

8. *Decretum Gratiani:* canon *Quoniam non cognovi*, Dist. LXXXVIII, canon 12; and canon *Qualitas lucri*, Dist. V, "de paenitentia," canon 2; *quia difficile est inter ementis vendentisque commercium non intervenire peccatum*. Cf. Schreiber, *op. cit.*, p. 129.

9. Jelle C. Riemersma, "Usury Restrictions in a Mercantile Economy," *Canadian Journal of Economics and Political Science*, XVIII (1952), 22.

1. See the encomium of trade by Thomas Mun, *England's Treasure by Forraign Trade* (London, 1664), chap. 21. Cf. Eli F. Heckscher, *Mercantilism* II, 281.

2. William D. Grampp, "The Liberal Elements in English Mercantilism," this *Journal*, LXVI (1952), 469. These ideas must have been current among the merchants on the continent as well as in England, since we find them also in Lodovico Guicciardini's famous description of Antwerp, first published in 1567: *Description de tous les Pays-Bas*, trans. François de Belleforest (Antwerp, 1582), p. 182; republished in *Tudor Economic Documents*, eds. R. H. Tawney and Eileen Power, III, 161.

with distrust, is extolled as the benefactor of humanity and the principal pillar of the State. This is what one might expect, since mercantilism was the economic system developed by, and for, the merchants.

In contrast to scholastic economics, mercantilism was amoral. The later mercantilists were interested in a large population and full employment only because they thought such conditions would stimulate trade and increase the economic power of the state.[3] Usury was no longer considered a voracious monster: Sir Josiah Child (1630–1699), Sir Thomas Culpeper the Elder, and others complained only that the interest rate, being higher in England than in Holland, favored the competition of the Dutch.[4] Trade has no soul and the individual did not count: why should the mercantilists be disturbed by moral issues?

One of the most striking characteristics of scholastic economics was universalism: regardless of origin and nationality, the Doctors are in fundamental agreement on method and principles. Although there may be, sometimes, sharp differences on points of detail or of practical application, all their treatises follow more or less the same pattern easily recognizable by anyone acquainted with scholastic literature. In the mercantilist camp, on the contrary, such uniformity in doctrine or method does not exist: neither between national schools nor between individual writers.

Among the mercantilists, "everyone is his own economist," according to the phrase so aptly coined by Professor E. A. J. Johnson. No one considers himself bound by precedent, and each author follows his own inspiration in selecting the appropriate method for dealing with his chosen topic.

Notwithstanding the great prestige of Eli F. Heckscher, I disagree with his statement that mercantilism strove toward unity.[5] As a matter of fact, non-scholastic economics in the seventeenth and eighteenth centuries varied greatly from country to country. In my opinion, the name "mercantilism" is appropriate only for British economics during that period. In Germany, one should speak of cameralism. One of its leading exponents, Johann Joachim Becher (1635–

3. E. A. J. Johnson, *Predecessors of Adam Smith*, pp. 247–52; Heckscher, *Mercantilism*, II, 159; Philip W. Buck, *The Politics of Mercantilism*, pp. 44–48, 65–66, 89–90.

4. Heckscher, *Mercantilism*, II, 286–89.

5. Heckscher, himself, in response to criticism of his book, was forced to concede that mercantilism failed as a unifying system: "Mercantilism," *Economic History Review*, VII (1936), 48. Cf. Herbert Heaton, "Heckscher on Mercantilism," *Journal of Political Economy*, XLV (1937), 374; J. F. Rees, "Mercantilism," *History*, New Series, XXIV (1939–1940), 130.

1682), "was still strongly influenced by the venerable Aristotelian tradition," albeit that he considerably modified the scholastic views.[6] In France, the expression "Colbertism," rather than "mercantilism," should be used to designate the economic policy of Colbert. Moreover, this policy aroused much criticism from writers such as Vauban (1633–1707) and his cousin Boisguilbert (1646–1714), whose comments on the iniquities of the French tax system anticipated the physiocrats instead of owing something to mercantilist ideas.[7]

Although the United Provinces were the leading economic power in the seventeenth century, there exists as yet no adequate study on Dutch economic thought during this period.[8] At any rate, Hugo Grotius or de Groot deserves a niche in the gallery of famous economists. One can hardly classify him as a mercantilist; he was rather an Aristotelian who used scholastic methods to defeat scholasticism.[9] Even Pieter de la Court (1618–1685), although not an Aristotelian, is far too liberal to pass for a mercantilist.[1]

In Spain, after 1600, economic writers, without breaking with scholasticism, were mainly concerned with the country's ailments: vellon inflation, vagrancy, depopulation, and economic decline. Whether this concern with pressing social and economic problems labels them as mercantilists remains a debatable point.[2] As in Spain,

6. de Roover, "Monopoly Theory prior to Adam Smith," *op. cit.*, p. 519. There is a new book on Becher by H. Hassinger, *Johann Joachim Becher (1635–1682): ein Beitrag zur Geschichte des Mercantilismus* (Vienna, 1951). The author apparently regards Becher as a mercantilist. Heckscher, however, states that the German cameralists "were imbued with a spirit of their own" (*Mercantilism*, II, 263).

7. *Ibid.*, II, 264. Cf. Hazel van Dyke Roberts, *Boisguilbert, Economist of the Reign of Louis XIV*, p. 255: "Boisguilbert had completely shaken off mercantilist thought."

8. The best study is still that of Etienne Laspeyres, but it is almost a century old: *Geschichte der wirtschaftlichen Anschauungen der Niederländer und ihrer Litteratur zur Zeit der Republik* (Preisschriften gekrönt und herausgegeben von der Fürstlich Jablonowski'schen Gesellschaft, Vol. XI, Leipzig, 1863).

9. de Roover, "Monopoly Theory prior to Adam Smith," *op. cit.*, pp. 521–22.

1. Heckscher (*Mercantilism*, I, 351) admits that the Dutch were "less affected by mercantilist tendencies than most other countries." His treatment of Dutch writers is based entirely on the study of Laspeyres (*op. cit.*, II, 263) and, moreover, is very superficial. See the pertinent remarks of Heaton (*op. cit.*, pp. 371 f.) about Heckscher's neglect of Dutch economic thought and policy in the seventeenth century.

2. They are mercantilists according to Earl J. Hamilton, "Spanish Mercantilism before 1700," *Facts and Factors in Economic History: Articles by former Students of Edwin Francis Gay*, pp. 214–39. This is an introductory survey which lists a few tracts and makes some general comments on the contents of the economic literature in Spain from about 1600 to 1700. The Latin treatises, including the important work of Luis de Molina, are not discussed. After stating that

so also in Italy the scholastic traditions were particularly strong, and persisted well into the eighteenth century along with other currents of thought originating in the merchant manuals of the Middle Ages.[3] In 1613, a Neapolitan writer, Dr. Antonio Serra, in fighting a scheme to regulate foreign exchange, formulated independently the balance-of-trade theory developed contemporaneously by the English mercantilists.[4] His proposals were dismissed, and his book was ignored for more than a century until abbé Ferdinando Galiani praised it as an outstanding performance. The witty abbé expresses his surprise that a book like Serra's was conceived "in an age of ignorance about economic matters," but he complains that the work is "tedious" reading because of its obscure style, its poor organization, and its "divisions and subdivisions" reminiscent of scholastic literature.[5] In other words, the abbé is a typical example of the eighteenth-century point of view. Another interesting fact is that Galiani considers the work of Serra to be scholastic, whereas most modern authors have classed it as a mercantilist pamphlet.[6]

The trouble is that the word "mercantilism" does not stand for a clear concept, but lends itself to confusion. The great specialist Heckscher, himself, has to admit that "mercantilism is simply a convenient term for summarizing a phase of economic policy and

most of the Spanish economic writers were ecclesiastics with no intimate knowledge of business or finance (pp. 229–30), Professor Hamilton calls them "mercantilists." Sancho de Moncada, one of the so-called Spanish mercantilists, was professor of theology in the University of Toledo, as Hamilton himself points out (*American Treasure*, p. 294). Other authors, including Andrés Villegas Castillo, Ramón Carande, Bernard W. Dempsey, Marjorie Grice-Hutchinson, and José Larraz, do not agree with Hamilton's classification. Only Gerónimo de Uztáriz (1670–1732), a late writer and statesman, seems to have come strongly under the influence of mercantilist thought. Cf. Ramón Carande, *Carlos V y sus Banqueros, la vida económica de España en una fase de su hegemonía, 1516–1556* (Madrid, 1943), p. 89.

3. Heckscher (*Mercantilism*, II, 263) implicitly admits that he is unacquainted with Italian economic literature. The famous tract of Bernardo Davanzati (1529–1606), *Notizia dei cambi*, written in 1581, was certainly based on merchant manuals, as appears clear from two manuscripts in the State Archives of Pisa: Fondo Alleati, Nos. 17 and 69. I owe this information to the kindness of Professor Federigo Melis of the University of Pisa.

4. *Breve trattato delle cause che possono far abbondare li regni d'oro e argento dove non sono miniere con applicazione al Regno di Napoli* in *Economisti del cinque e seicento*, ed. Augusto Graziani (Bari, 1913), 141–233. Selections from Serra's treatise, in English translation, are found in Monroe, *op. cit.*, pp. 143–67.

5. Ferdinando Galiani, *Della moneta* (Bari, 1915), p. 344.

6. Monroe, *op. cit.*, p. 144; Cossa, *op. cit.*, p. 178; Fanfani, *Storia, il volontarismo*, p. 171; Lewis H. Haney, *History of Economic Thought* (3d ed.), pp. 112–13; John M. Ferguson, *Landmarks of Economic Thought*, pp. 36–37.

economic ideas."[7] It should be added that the term covers only those heterogeneous ideas that are non-scholastic in inspiration.

There are remnants of scholastic influence in many mercantilist writings, but surprisingly those traces have not been recognized, though they are not so difficult to spot. The mercantilists, of course, were unable to escape from the impact of several centuries of culture. Whether or not they knew it, they absorbed some of the ideas bequeathed by former generations.[8]

Gerard de Malynes is the writer in whose works the traditional views are the most perceptible. Whether he should be considered as a mercantilist or as a scholastic writer, is to my mind a moot question.[9] In any case, there can be no doubt that he forms the link between the two schools of thought. His insistence on the par as the only fair rate of exchange is simply a variant of the just price theory taken over from Dr. Thomas Wilson, himself a Doctor still imbued with scholastic traditions. According to Professor Jacob Viner, Malynes was poor in market analysis,[1] but there can be no question about his being well read and well acquainted with ancient and scholastic literature.[2] In his *Saint George for England*, a tract against usury, he describes the dragon called *Foenus politicum* as having two wings, *usura palliata* and *usura explicata*, and a tail, "inconstant *Cambium*."[3] This allegory is obviously sheer and unadulterated scholasticism. Malynes has also received credit for distinguishing between changes in the price level due to monetary factors and changes in the price of particular commodities due to the operation of the law of supply and demand. I strongly suspect that this idea did not originate with him but that he took it from a continental treatise, for he was by no means an original thinker and was addicted to plagiarism.[4]

7. "Mercantilism," *Economic History Review*, VII (1936–37), 54.

8. Heckscher (*Mercantilism*, II, 277) states: "Here one may perceive a tendency towards economic liberty that was never entirely broken off and therefore connected medieval and laizzez-faire ideals."

9. de Roover, *Gresham on Foreign Exchange*, pp. 285 f.

1. *Studies in the Theory of International Trade* (New York, 1937), p. 76.

2. Helen E. Sandison, "An Elizabethan Economist's Method of Literary Composition," *Huntington Library Quarterly*, VI (1942–43), 205–11. Professor Sandison shows that Malynes certainly "borrowed" from Sir Thomas More's *Utopia*. I may add that he also was acquainted with the works of Jean Bodin, Lodovico Guicciardini, Dr. Thomas Wilson, Aristotle, and most probably, Leonardus Lessius.

3. *Saint George for England allegorically described* (London, 1601); "Foreword to the Reader." On p. 61, Malynes mentions the extrinsic titles, *damnum emergens* and *lucrum cessans*.

4. In the sixteenth century, most of the scholastic writers accepted the quantity theory of money and stated that prices "generally" go up or down

In a recent article, the mercantilists have been praised for the "liberalism" of their concepts.[5] Contrary to the conclusions of the author, it appears, however, that those so-called "liberal elements" are rooted in the doctrines of the medieval Schoolmen.[6] For one thing, the Doctors were uncompromising in their condemnation of monopoly for the reason that the monopolist exploits the public and makes an illicit gain by raising the price of his articles above the competitive level. For example, Cardinal Cajetan, commenting on the *Summa* of Thomas Aquinas states that monopoly offends freedom by compelling the public to pay a price higher than the one that would prevail in the market, if there were no such monopoly (*si huiusmodi monopolium non esset*).[7] The traditional feeling against monopoly was so strong that no mercantilist writer dared openly defy public opinion, even when his purpose was to justify the monopolistic practices of this or that trading company.[8] In the parlance of the mercantilists, "free trade," as I have pointed out in this *Journal*, meant freedom from restraints of any sort in internal as well as in foreign trade. Consequently, it corresponded to the French expression *liberté du commerce* and not to *libre échange*.[9] In the seventeenth century, protection in the modern sense was not yet born; the struggle was still a medieval struggle for the control of the carrying trade.[1] In dealing with the history of economic thought, it is not enough to know the writings of the economists; one must also know something

with the abundance or scarcity of money. Such a statement had even become commonplace.

5. Grampp, *op. cit.*, pp. 465–501.

6. *Ibid.*, pp. 500 f. So far as I know, the Schoolmen have never stated "that free individual behavior was inimical to the welfare of society." Heckscher (*Mercantilism*, II, 277) asserts the contrary and rightly states: "that even the medieval tradition was sympathetic to a certain sort of freedom. The medieval influence was thus not without importance to the notion of economic liberty under mercantilism." As late as the seventeenth century, the Anglican and Puritan divines continued to propound scholastic doctrine on just price, monopoly, and price discrimination. See the characteristic passages of Richard Baxter (1615–1691), a popular preacher, which are quoted by H. M. Robertson, *Aspects of the Rise of Economic Individualism: A Criticism of Max Weber and his School* (Cambridge, 1935), p. 17.

7. Text quoted by Fanfani, *Origini dello spirito capitalistico*, p. 123. Cf. Joseph Höffner, *Wirtschaftsethik und Monopole im fünfzehnten und sechzehnten Jahrhundert* (Jena, 1941), p. 107.

8. de Roover, *Gresham on Foreign Exchange*, p. 284. Such was certainly the purpose of John Wheeler, Edward Misselden, Thomas Mun, Sir Josiah Child, and Charles Davenant.

9. When French authors of the period mean *libre échange*, they use the expression: *liberté du commerce entre les nations*.

1. de Roover, *Gresham on Foreign Exchange*, pp. 282 f.

about the institutional framework and the social environment of the period.

Certainly, the English "mercantilists did not believe in an economy wholly or mainly directed by the State,"[2] but they wanted the state to pursue a policy favorable to the trading interests and they tended to defend the exclusive privileges of chartered companies and corporations.[3] Owing to the persistent influence of scholastic ideals, the mercantilists paid lip service to the goddess of "free trade," though the sincerity of their devotion is very much open to question, inasmuch as their pretenses conflict with their other aims. But then, mercantilism was not a logical system. It may even plausibly be argued that, unlike scholasticism, the much vaunted mercantile system was not a system at all.

V. Conclusions

The shortcomings of scholastic economics — and no effort has been made to conceal them — should not blind us to the greatness of the achievement. The Doctors correctly diagnosed the economic problem as one of scarcity. In their opinion, economics was a branch of ethics which determined the rules of justice that ought to preside over the distribution and the exchange of scarce goods. It is obvious that there would be no need for distribution or exchange, if goods could be obtained without effort in unlimited quantities.

The great difference between scholastic and contemporary economics is one of scope and methodology: the Doctors approached economics from a legal point of view. They attached an excessive importance to formalism, so that the study of economics nearly reduced itself to an investigation into the form and nature of contracts. Because of their preoccupation with ethics, the Doctors were also more interested in what ought to be than in what actually was. In the matter of usury, they made the fatal mistake of allowing this subordinate question to overshadow all other problems. Besides, the sophistication of the later casuists involved them more and more in a maze of contradictions, which, ever since the eighteenth century, have prejudiced economists against scholastic doctrines. The more concessions the casuists made, the more they undermined their own position. They were unwilling to face the fact that their distinction

2. Grampp, *op. cit.*, p. 495.
3. In order to enlist the support of the government, mercantilist writers and projectors never failed to stress the benefits which would accrue to the Royal Treasury, if their schemes were carried out (Heaton, "Heckscher on Mercantilism," *op. cit.*, p. 376).

between usurious and nonusurious contracts was based on mere legal technicalities.[4]   After all, was it logical to allow a charge for the use of money in one case and to prohibit it in another?[5]

The Doctors, especially the members of the school of Salamanca, made one of their main contributions in developing a theory of value, based on utility and scarcity, which is more in line with modern thinking than that of Adam Smith.   Because of his influence and prestige, he created a century of confusion on this topic by throwing out utility and by becoming entangled in the antithesis of value in use and value in exchange.   The Doctors were also right in stressing from the beginning the principle of mutual advantage in any bargain or voluntary exchange.[6]

In the absence of fraud or collusion, the market price was supposed to be just, but the Doctors never questioned the right of public authorities to interfere, whenever, because of famine or other circumstances, either buyers or sellers would be seriously harmed by the free operation of the law of supply and demand.   Perhaps in the nineteenth century, economists might have regarded the scholastic position as erroneous, but today we operate, in fact, on a just-price basis, since the government does not hesitate to regulate prices in times of national emergency.   While the Doctors may have been correct in their analysis, they had the fault of many idealists of overlooking entirely practical difficulties: they assumed that it sufficed to set a price by decree in order to make it effective.

In accordance with the teachings of the Doctors, monopoly was almost everywhere considered a criminal offense.   Incidentally, the Doctors rarely mention the guilds and then only to reprove them for their monopolistic practices.[7]   I do not find evidence in their treatises that they favored the guild system, which is so often pictured as an ideal organization for Christian society or is recommended as a panacea against the evils of modern industrialism.[8]

4. Robertson, *op. cit.*, p. 118: "In practice, Calvin's position (that all usuries were not necessarily to be condemned) had been reached by Catholic teachers. The difference was mainly one of expression.   Amongst the Catholics more depended upon the formalities of contracts."

5. See R. H. Harrod's remarks in a review of John P. Kelly, "Aquinas and Modern Practices of Interest Taking," *Economic Journal*, LVI (1946), 314.

6. Grampp (*op. cit.*, p. 466) credits the mercantilists with the formulation of this principle, although it is clearly stated by Thomas Aquinas, in his *Summa theologica*, II, ii, qu. 77, art. 1, corpus.   Cf. Monroe, *op. cit.*, p. 54.   John Buridan gives an even better analysis than Aquinas (Schreiber, *op. cit.*, p. 183).

7. Höffner, *Wirtschaftsethik und Monopole*, pp. 82, 92–94.

8. This is the point of view, for example, of the advocates of guild socialism, especially the earlier adherents of this school.   Cf. Arthur J. Penty, *Old Worlds*

Among other contributions of the Doctors, one should not forget to mention their acceptance, by the sixteenth century, of the quantity theory of money and their speculations on the lawfulness of banking and dealings in foreign exchange. The latter discussion, starting as far back as the thirteenth century, paved the way to the balance-of-trade theory, developed by the English mercantilists in the Tudor and Stuart period. Unfortunately, the late casuists never paid any attention to this discovery and even allowed it to be used against them by their opponents.

One should not mistakenly assume that scholastic economics exerted no influence on business morality. The Church sought to enforce its code of social ethics in two ways: *in foro externo*, that is, through the courts, ecclesiastical and secular, and *in foro interno*, that is, through the confessional. In the Middle Ages, all over western Europe, usurers were constantly brought to court. It is true that the historian stops at the threshold of the confessional, but the numerous medieval wills providing for restitution of usury suggest that confession was far from being an ineffective means of enforcement.[9]

This is not the place to enter into a discussion of the Max Weber theory about the rôle of religion in the rise of capitalism. I question it, because the writings of the Doctors seem to show that the medieval Church neither favored nor hindered the development of capitalism. Like technological and scientific progress, capitalism grew outside the Church. It does not follow, however, that scholastic doctrines had no influence on the course of economic development. Quite the contrary. Recent research in economic history has established the fact that the usury prohibition profoundly affected the development of banking. Since the taking of interest was forbidden, the discounting of commercial paper was also ruled out, but the bankers cleverly shifted to exchange dealings as the basis of their operations. This shift changed the entire structure of the continental European banking system up to the time of the French Revolution.[1]

To consider scholastic economics as medieval doctrine is simply an error, and economists have bypassed a current of thought which

for *New, a Study of the Post-Industrial State*, pp. 44–49; Ralph Adams Cram, *Walled Towns* pp. 46, 80–82; G. D. H. Cole, "Guild Socialism," *The Encyclopaedia of the Social Sciences*, VII, 202–4.

9. Benjamin N. Nelson, "The Usurer and the Merchant Prince: Italian Businessmen and the Ecclesiastical Law of Restitution, 1100–1550," Supplement to *The Journal of Economic History*, VII (1947), 104–22.

1. de Roover, *L'évolution*, pp. 144–45.

runs parallel with mercantilism and reached out into the eighteenth century, connecting the *économistes* and even Adam Smith with Thomas Aquinas and the medieval Schoolmen.[2] Traces of scholastic influence still permeate eighteenth century economic thinking and sometimes appear in unexpected places, such as the *Encyclopédie* of Diderot and d'Alembert. The *Encyclopédie's* definition of price differs in no way from that given in scholastic treatises, and the same applies to the treatment of monopoly and dry exchange.[3]

In the case of Adam Smith, the ascendance which links him to scholasticism passes through his teacher, Francis Hutcheson, Samuel Pufendorf, and Hugo Grotius.[4] Smith's library contained copies of both Grotius and Pufendorf.[5] Moreover, there is evidence that Adam Smith read Grotius at the age of fifteen when he was a student at Glasgow College. At that time, his teacher was using as textbook a translation of Pufendorf's *De officio Hominis et Civis* by Gershom Carmichael (d. 1729), Hutcheson's predecessor in the chair of Moral Philosophy.[6] In his lectures on political economy, as already stated, Hutcheson dealt with the subject in scholastic fashion as a branch of natural jurisprudence, particularly as "a discussion of contracts."[7] When Adam Smith, himself, succeeded to the chair of Moral Philosophy, he modified this outline by transferring economics to the fourth part of "his course of lectures" devoted to matters not pertaining to justice, but to expediency.[8] This decision definitely constituted a break with the scholastic tradition. The outline of the course in Moral Philosophy, as taught by Francis Hutcheson and later by Adam Smith himself, clearly shows that the curriculum of Glasgow College, in the eighteenth century, never paid any attention to mer-

2. Professor Mabel Magee, my former colleague at Wells College, tells me that Seligman was an exception. According to her detailed notes on his course at Columbia University on the history of economic thought, he dealt with most of the writers mentioned in this article and did not consider scholastic economics as a medieval economic doctrine. I avail myself of this opportunity to thank Dr. Magee and another former colleague, Professor Jean S. Davis, for reading a draft of this article and making helpful suggestions.

3. According to the *Encyclopédie*, the price of commodities is set either by ordinance or by common estimation: the first is called the legal price (*prix légitime*) and the second, the current price (*prix courant*). The scholastic origin of this distinction is beyond question.

4. Grice-Hutchinson, *The School of Salamanca*, pp. 64–69, 76. Cf. William Robert Scott, *Adam Smith as Student and Professor* (Glasgow, 1937).

5. James Bonar, *A Catalogue of the Library of Adam Smith* (2d ed.), pp. 78, 151.

6. Scott, *op. cit.*, pp. 34, 112.

7. John Rae, *Life of Adam Smith*, p. 14.

8. *Ibid.*, pp. 54 f.

cantilist thought, but always provided for some teaching of economic principles based on ethics and law, inherited from the medieval university. In the *Wealth of Nations*, Adam Smith, it is true, devotes several chapters to mercantilism but only to denounce it as a pernicious and "sophistical" system.

Two eighteenth century economists, abbé Ferdinando Galiani (1728–87) and abbé Etienne Bonnot de Condillac (1715–80), have been hailed by some historians as the first to anticipate the modern marginal utility theory of value by stating that value rests on the combination of two elements: utility and scarcity.[9] The question arises whether this idea originates with the two abbés or whether — what is more likely — they took it from the Doctors, possibly by way of the late casuists and the school of Salamanca, as Marjorie Grice-Hutchinson seems to think.[1] In my opinion, she is undoubtedly right, since it is highly improbable that cultured men in holy orders would be unacquainted with the extensive literature on moral theology. As far as Galiani is concerned, scholastic influence is noticeable in many passages of his essay on money, especially in his treatment of usury and *cambio*.[2] Furthermore, the chapter on value contains a quotation from Diego Covarrubias y Leyva, one of the leading representatives of the school of Salamanca.[3] Consequently, Galiani certainly knew his work, and hence there was no breach of continuity.

As this study shows, modern economics owes the Schoolmen and their successors a greater debt than is commonly acknowledged. It also illustrates the merits and the defects of the legal approach to economics. What really caused the downfall of scholastic economics was the refusal of the late casuists to revise and to modernize their methods. Perhaps their whole system was in need of a complete overhauling. Nevertheless, it contained much that was worth pre-

9. Galiani, *op. cit.*, Book I, chap. 2, pp. 25–45; and Condillac, *op. cit.*, Vol. I, chaps. 1 and 2, pp. 248–57. In a footnote, the editor, Eugène Daire, blames Condillac for not following in the footsteps of Quesnay and Adam Smith and not adopting their distinction between *value in use* and *value in exchange!*

1. *Op. cit.*, pp. 63–64, 76.

2. *Op. cit.*, Book 5, chaps. 1 and 4, pp. 289–96, 303–7. Galiani's definition of usury, as any gain above the principal accruing from a *mutuum*, is still purely scholastic.

3. *Op. cit.*, Book I, chap. 2, p. 26. The most recent and much the best biographical study on abbé Galiani is that of President Luigi Einaudi, "Galiani economista," *Saggi bibliografici e storici intorno alle dottrine economiche* (Rome, 1953), 269–305. This study was first published in German under the title "Galiani als Nationalökonom," *Schweizerische Zeitschrift für Volkswirtschaft und Statistik,* LXXXI (1945), No. 1. An English version of the first part of Einaudi's study is available in Henry William Spiegel (ed.), *The Development of Economic Thought* (New York, 1952), pp. 62–82.

serving and which was preserved in actual fact. Valuable ideas may lie buried for a time but they eventually spring up. Like other sciences, economics grows slowly by accretion. Despite many currents and cross currents, continuity is perhaps the most impressive phenomenon in the history of economic doctrines.

<div align="right">RAYMOND DE ROOVER.</div>

THE GRADUATE SCHOOL
BOSTON COLLEGE

ACCELERATED DEPRECIATION ALLOWANCES AS A STIMULUS TO INVESTMENT*

By Richard Goode

Over the past fifteen years considerable attention has been given to the possibility of influencing the volume of private investment by adjusting allowances for depreciation or amortization of capital assets under income and profits taxes. During World War II and again after the outbreak of hostilities in Korea, the United States increased normal depreciation or amortization allowances for approved defense facilities. In the postwar period the United Kingdom, Australia, Canada, India, Pakistan, Sweden, the Netherlands, and other countries have liberalized depreciation allowances with the objective, at least in part, of encouraging ordinary private investment. Most recently, interest in the subject has been revived by the provision for liberalized depreciation allowances in the comprehensive revision of the tax laws adopted in the United States in 1954 and by introduction in the U.K. budget for 1954–55 of a new "investment allowance" intended to offer a greater stimulus to re-equipment of industry than was provided by the previous "initial allowance."

The present paper presents an analysis of accelerated allowances for depreciation as a stimulus to investment and briefly compares this approach with a reduction in tax rates and a reduction in interest

rates.   As here used, the term "accelerated depreciation allowances" — or, in brief, "accelerated depreciation" — covers all methods of deliberately speeding the rate at which the original cost of assets (less any salvage value) may be deducted from taxable income.   Part I of the paper deals with the technique in general but contains a few references to particular schemes for the purpose of illustrating general principles.   Part II compares different methods of accelerating depreciation allowances.[1]

I. General Discussion

A. *Normal and Accelerated Depreciation Allowances*

Any tax on net income or profits must make some allowance for depreciation, that is, for amortization of the cost of durable plant and equipment.   Several methods are employed for apportioning these allowances.   The two most common are: (1) the straight-line method, under which the cost of depreciable property is written off in equal annual installments over its expected useful life; and (2) the declining-balance method, under which the deduction is a constant fraction of the unamortized balance of the cost of the asset but, of course, a declining fraction of original cost.[2]   The straight-line method has been commonly employed in the United States, whereas the declining-balance method has been more usual in the British Commonwealth. Separate allowances are sometimes made for wear and tear and for obsolescence, but in the present discussion no distinction will be made between these two causes of exhaustion of usefulness.

Although timing of normal depreciation deductions differs significantly under the straight-line and the declining-balance methods, both systems spread the allowances over the whole of the expected useful life of the asset.   Indeed, aggregate deductions under the

---

1. Analyses of certain aspects of accelerated depreciation can be found in: E. Cary Brown, "Business-Income Taxation and Investment Incentives," in *Income, Employment and Public Policy: Essays in Honor of Alvin H. Hansen*, pp. 300–16; Evsey D. Domar, "The Case for Accelerated Depreciation," this *Journal*, LXVII (Feb. 1953), 493–519; S. P. Dobrovolsky, "Depreciation Policies and Investment Decisions," *American Economic Review*, XLI (Dec. 1951), 906–14; Robert Eisner, "Accelerated Amortization, Growth, and Net Profits," this *Journal*, LXVI (Nov. 1952), 533–44; Richard E. Slitor, "Liberalization of Depreciation," *Proceedings of the National Tax Association*, 1953, pp. 466–74.   See also Domar, "Depreciation, Replacement and Growth," *Economic Journal*, LXIII (Mar. 1953), 1–52; Eric Schiff, "A Note on Depreciation, Replacement, and Growth," *Review of Economics and Statistics*, XXXVI (Feb. 1954), 47–56.

2. For an outline of these and other, less widely used, methods, see *Accountants' Handbook*, ed. W. A. Paton, 3d ed., pp. 751–65.

declining-balance method equal total original cost only if a final adjustment is made when the asset is retired.

Accelerated depreciation implies a concentration of depreciation allowances in the early years of an asset's life. At the limit, outlays for durable capital goods may be deducted from current income like ordinary operating expenses, but most plans are less extreme. The objective of accelerated depreciation may be partly to approximate more closely than established accounting procedures do the supposed loss in economic value of the asset, or it may be to offer a tax advantage to new investment by allowing during the period of acceleration deductions in excess of capital costs as measured by the actual loss in economic value or by conventional cost-allocation formulas. In either case, liberalization of depreciation allowances may act as a stimulus to investment. Accelerated depreciation has also been recommended as a means of relieving financial problems which arise in a period of inflation because replacement costs of depreciable property exceed historical costs,[3] but this aspect of the subject will not be considered in the present paper.

The simplest method of accelerating depreciation allowances would be merely to apply some multiplier to normal allowances (with the proviso that allowances cease after the full original cost has been recovered). In practice, however, other techniques have been more popular. The United Kingdom, for example, has granted investors an "initial allowance" in addition to normal depreciation allowances, which can be deducted from taxable income in the year in which an eligible asset is acquired. This initial allowance is only an anticipation of normal amortization and correspondingly reduces the deductions allowable in later years. On the other hand, the "investment allowance," introduced in the 1954-55 budget, is a net addition to allowable deductions and not merely an acceleration of allowances previously granted. The United States has permitted taxpayers to write off the full cost of emergency defense facilities over a five-year period, by the straight-line method, even when the normal useful life exceeds five years. The recent U.S. revenue act offers taxpayers the option of using for ordinary investment a combination of the declining-balance method and the straight-line method which will permit a significant increase in deductions during the early years

3. United Kingdom, *Report of the Committee on the Taxation of Trading Profits* (London, 1951, Cmd. 8189), pp. 35–46. For critical comments, see E. Cary Brown, *Depreciation Adjustments for Price Changes* (Boston, 1952), pp. 115–23.

of an asset's life. These methods are discussed in greater detail in Part II.

B. *Influence on Investment*

In appraising the influence of accelerated depreciation it will be convenient at the outset to abstract from changes in tax rates and business conditions and to assume that the taxpayer always has enough taxable income to absorb any increase in deductions allowed under accelerated depreciation. These assumptions will be relaxed at a later stage of the analysis.

Time Discount. The first and most obvious point to be noted about accelerated depreciation is that the taxpayer realizes an interest or time-discount gain by receiving the tax benefit of depreciation allowances at an earlier date. The simplest illustration of this point is provided by reference to five-year amortization for certain emergency facilities in the United States. It will be recalled that this plan allows straight-line amortization over five years of the cost of eligible assets. If the marginal tax rate applicable to the investor throughout the life of the asset is 50 per cent, the reduction in tax payments attributable to deduction from taxable income of depreciation charges with respect to an item costing $100 will total $50 by the end of the asset's useful life. If money is worth 6 per cent per annum to the taxpayer, the present value of this $50 tax saving is $42.12 if the deductions are distributed over five years by the straight-line method.[4] The present value of the tax saving from the same total amount of deductions distributed over ten years would be $36.80; over twenty years, $28.67; over forty years, $18.81. Thus, the excess of the present value of the tax saving from five-year amortization over that from normal straight-line depreciation amounts to 5.32 per cent of the original cost of a ten-year asset, 13.45 per cent for a twenty-year asset, and 23.31 per cent for a forty-year asset. These margins vary directly with the length of the normal life of the asset and the marginal tax rate. Over a wide range of interest rates, the margins also vary directly with the interest rate, but, for any given length of normal life, time-discount gain reaches a maximum at some (high) interest rate and thereafter declines.[5]

4. Assuming, for the sake of convenience in this and the following illustrations, that the tax savings are realized in a series of annuities at the end of each year. This simplification does not materially affect the results.

5. The maximum will occur at lower discount rates for long-lived than for short-lived assets. For example, the maximum time-discount gain from five-year amortization would be realized at a discount rate of around 11 per cent for forty-

Exact computations of the present value of future deductions from taxable income, as implied in the foregoing illustrations, may be exceptional in actual business practice. It seems reasonable to believe, however, that virtually all businessmen recognize an advantage in receiving income in the near future as compared with the same amount of income in the more distant future, even though they may not make a formal calculation of the present value of the two sums.

The greater present value of the near-term tax saving made possible by accelerated depreciation is in many respects similar to a reduction in cost of the capital asset. Under any given set of operating and market conditions this improves the net return on the investment. But the quantities involved may be too small to have a significant influence on investment unless interest rates and tax rates are high.

Risk or Uncertainty. A second and often a much more important way in which accelerated depreciation may stimulate investment is through its influence on risk or uncertainty. In combination with normal depreciation allowances, an income or profits tax may seriously interfere with investors' plans for protecting themselves against risk or uncertainty. Accelerated depreciation will greatly reduce this hazard.

The available fragmentary information on American and British practices indicates that, insofar as businessmen make any systematic allowance for risk and uncertainty in investment decisions, they most often do so by insisting that a new asset "pay for itself" in a time period considerably shorter than its normal physical or economic life. The "pay-off period" is that time interval over which use of the new asset is expected to reduce operating expenses or to increase net profits (before depreciation allowance) by an amount equal to the cost of the asset. "Cost" sometimes includes an interest allowance but often does not. For machinery of a type having a normal useful life of ten to twenty years the "pay-off period" may be as short as two or three years and is reported to be seldom longer than five years.[6]

year assets and at around 25 per cent for ten-year assets. I had originally overlooked the fact that time-discount gain reaches a maximum but discovered the point in a paper by E. Cary Brown, "1954 Changes in Depreciation under the Income Tax," to be published in the *National Tax Journal.*

6. Ruth Mack, *The Flow of Business Funds and Consumer Purchasing Power,* pp. 255–59; George Terborgh, *Dynamic Equipment Policy,* chap. XII, pp. 187–201; P. W. S. Andrews and Elizabeth Brunner, *Capital Development in Steel* (Oxford, 1951), chaps. VIII and X, pp. 252–300, 347–67.

Despite its crudity, the pay-off-period approach seems to be a very sensible method of allowing for risk and uncertainty. Literally interpreted, it implies that uncertainty regarding the future is so great that the possibility of realizing earnings from the asset after the end of the pay-off period is not worth taking into account. Although this is an extreme assumption, it is clearly true that risk increases with the duration of a commitment and that uncertainties of the distant future are much greater than those of the near future.[7]

For investors who employ a short pay-off requirement as a method of allowing for risk or uncertainty, the relation between their subjective pay-off period and the normal depreciation period for tax purposes is highly important. If an investor's required pay-off period is shorter than the normal depreciation period, as it typically is, his amortization requirements over the pay-off period will exceed allowable depreciation deductions. This excess will be subject to income or profits tax, and net "pay-off" realizations will be reduced. Hence it will be impossible fully to amortize some marginal projects out of their net yield within the investor's planning period. If, however, the depreciation period for tax purposes is as short as the pay-off period, the income tax will not interfere with plans for amortization of an investment.

The significance of accelerated depreciation with a short pay-off period may be illustrated by again referring to the five-year amortization plan which has been in effect in the United States. This plan allows a tax-free return up to 20 per cent a year for the first five years. Projects yielding a little more than 20 per cent a year, before taxes and depreciation allowances, may therefore be made acceptable to an investor who insists on a five-year pay-off period, even though they have a normal useful life of ten, twenty, or more years. On the other hand, with normal straight-line depreciation and an income tax rate of 50 per cent, ten-year assets must yield at least 30 per cent to satisfy a five-year pay-off requirement and twenty-year assets must yield at least 35 per cent.[8] Introduction of five-year amortization, there-

---

7. For some investors a short pay-off-period requirement may be merely a method of rationing a limited supply of capital among a wide range of profitable opportunities. These investors cannot be expected to respond to accelerated depreciation to the same extent as those who insist on a short pay-off period as an allowance for risk. But investors whose primary concern is to choose among projects which are adequately profitable, at the existing tax rate, do not need a tax stimulus. If such investors should predominate at any time, it will not be socially necessary to raise the subjective rate of return on investment, although it may be desirable to make capital more readily available.

8. With normal straight-line depreciation the tax-free return on a ten-year asset is 10 per cent and on a twenty-year asset 5 per cent. With a 50 per cent tax

fore, will bring within the range of favorable consideration ten-year assets which are expected to yield between 20 and 30 per cent (before taxes) and twenty-year assets expected to yield between 20 and 35 per cent. [9] This, of course, does not mean that all projects within the range will immediately be accepted. There may be enough projects offering still higher yields to use up the capital and managerial resources available to the investor.

These illustrative computations are based on the assumption of strict annual accounting for tax purposes, that is, no allowance is made for the possibility that a net loss suffered after the end of the pay-off period may result in a refund of taxes paid in earlier years. The possibility of tax refunds due to a carryback of net operating losses, however, may considerably lessen the damage done by a depreciation period for tax purposes which is longer than the investor's pay-off period. Accounting losses will be suffered with respect to a particular project after the end of its pay-off period if in practice it conforms to the conservative assumption that it will yield no return at that time. These accounting losses will equal normal depreciation deductions on the asset or (under most income tax laws) the undepreciated balance of cost, less any salvage value, if the asset is discarded. With an unlimited carryback, these losses could be deducted against any taxable profits earned during the pay-off period, and the investor would be entitled to a refund of taxes previously paid. If, on the other hand, the asset continued to yield a profit after the end of its pay-off period, the investor's original planning would be proved to be unduly pessimistic. In either case, the income tax would not prevent amortization of the investment out of earnings, although full amortization would be delayed.

In practice, the period over which losses may be carried back against prior income is usually severely limited. For example, in the United States the carryback has only recently been extended from one year to two years. Even a short carryback may be highly significant, however, when it covers a substantial fraction of a brief pay-off period. Thus, with a two-year carryback and a 50 per cent tax rate, the discounted value of the gross annual yield necessary to satisfy a five-year pay-off requirement on a ten-year asset which is subject to straight-line depreciation for tax purposes is 26 per cent or 22.5 per

the investor can realize a 20 per cent return (gross of depreciation but net of tax) only by achieving an additional before-tax return of 20 per cent on the ten-year asset and 30 per cent on the twenty-year asset.

9. These yield figures should be understood to be net of allowance for time discount.

cent, depending on whether the investor assumes that the asset will remain in use after the end of the fifth year, but produce no income, or will be discarded at the end of the fifth year with zero salvage value.[1] These figures contrast with the 30 per cent gross return which would be necessary in the absence of a carryback. The effect of a carryback is, of course, less marked when it covers a smaller fraction of the pay-off period than it does in this illustration.[2] It is apparent, moreover, that the availability of carryback refunds depends on technical features of the tax law and on actions of the taxpayer which may appear to be inconsequential in other respects.

Despite its potential importance, a carryback of losses will never completely neutralize the disincentive effects of a long depreciation period. Even if an unlimited carryback were provided by law, some uncertainty would surround its operations, merely because they lie in the future. The taxpayer would have to recognize the possibility that the law would be modified in the future and the possibility that he will find it advantageous or necessary to go out of business after the end of the pay-off period and hence will not be eligible for carrybacks. Eligibility for carryback refunds will often depend on conditions existing after the end of the investor's normal planning period.

It should be emphasized in this connection that a loss carryforward, as distinguished from a carryback, does not compensate for a discrepancy between the normal depreciation period and the pay-off period. Although a carryforward may be helpful in averaging out irregularities of taxable income within the pay-off period, it does not permit normal depreciation allowances accruing after the end of the pay-off period to be charged back against income realized during that period. Hence, apart from its averaging feature, a carryforward does not counteract the adverse effect of the income tax on the opportunity

1. If the asset remains in use but produces no income, the taxpayer will be entitled to carry back from year 6 and from year 7 net operating losses equal to normal depreciation allowances, in this case, 10 per cent of original cost each year. If, however, the asset is discarded with zero salvage value, the unrecovered balance of cost (50 per cent) will be charged against income of year 5, which will wipe out the tax liability of that year and produce a net loss (32.5·per cent) to be carried back against taxable income of years 3 and 4. The lower effective rate of taxation in this case will mean that a smaller return (in years 1–5) will be needed to amortize the asset than will be necessary if it is retained in a standby capacity.

2. For example, with a one-year carryback the range of yields necessary for a five-year pay-off on a ten-year asset is 24.25–28 per cent. To meet a ten-year pay-off period on a twenty-year asset (with a 50 per cent tax rate and normal straight-line depreciation) the minimum annual yield is 15 per cent with no loss carryback and 12.7–14 per cent with a two-year carryback. All figures mentioned in this and the preceding footnote and in the text paragraph should be understood as net of time discount.

of amortizing a capital asset out of the limited series of annual yields that fall within the investor's planning horizon.[3]

The argument presented in the preceding paragraphs concerns the influence of the income tax on the possibility of amortizing an investment out of expected earnings from that particular asset. To put the matter another way, the reasoning implies that investors who plan on the basis of a short pay-off period not only assume that the project under consideration will yield no return after the end of its pay-off period, but also disregard the possibility of realizing taxable income from other sources at that time. If, however, an investor has enough taxable income from other sources, normal depreciation deductions on an asset may continue to result in a tax saving after the asset itself ceases to yield a return. An investor who expects to be in this position may therefore consider a long depreciation period for tax purposes as merely postponing amortization of an asset with a short pay-off period rather than denying full amortization. The expectation that the firm will enjoy taxable income over the indefinite future is not necessarily inconsistent with adoption of a short pay-off period requirement for each individual project inasmuch as the investor may regard this requirement as a safeguard against loss under unfavorable conditions rather than a forecast of average experience. Especially when individual projects are small in relation to the total capital of the firm and are diversified in nature, the investor may feel justified in being more optimistic about the future of the enterprise as a whole than about any one project.

Too little is known about the attitudes of investors to reveal the extent to which they count on completing amortization of economically short-lived assets by deducting a part of their cost from taxable income to be realized from other sources in the future. The safest conclusion seems to be to concede that this possibility does reduce the potential contribution of accelerated depreciation but at the same time to recognize that it is always less satisfactory to the investor than full amortization within the pay-off period in view of the increased degree of risk and the time discount applicable to future income.

It must be recognized that pay-off periods seldom are as inflexible as implied by the foregoing discussion. Nevertheless, once it is granted that some investors allow for risk and uncertainty by insist-

3. S. P. Dobrovolsky (*op. cit.*, pp. 906–14) correctly stresses the importance of income averaging in connection with the depreciation problem but does not emphasize the difference between carrybacks and carryforwards.

ing on a pay-off period that is shorter than the normal depreciation period, accelerated depreciation suggests itself as a means of reducing the deterrent influence of taxation on new investment. A full analysis of the economic effects of the income tax would have to consider the possibility that investors' attitudes toward the appropriate pay-off period and interest rate are subject to alteration through changes in taxation. These questions, however, are outside the scope of the present paper.

Investors who allow for risk by applying a high rate of discount to future returns, rather than by adopting a short pay-off period, will also be favorably influenced by accelerated depreciation.[4] For example, application of a uniform 15 per cent annual rate of discount for risk in connection with a ten-year investment will reduce the present value of normal straight-line depreciation allowances over the life of the asset to roughly 50 per cent of original cost. Introduction of five-year amortization will raise the present value of depreciation charges to approximately two-thirds of original cost.

Availability of Funds. A third way in which accelerated depreciation may stimulate investment is by increasing the availability of financial resources. In view of imperfections of the capital market and the reluctance of many businessmen to seek outside capital, this consideration is not adequately reflected in the interest discount factor. When tax rates are high, a growing firm will be able to finance a substantially larger fraction of its investment from retained profits under an accelerated depreciation plan than with normal depreciation allowances for tax purposes. Although accelerated depreciation for any particular asset represents only an advance of tax savings which would be realized in the future with normal depreciation, this advance does not have to be repaid unless the firm reduces its rate of acquisition of new depreciable property for purposes of replacement and expansion.

Domar has explored this aspect of the subject in his illuminating article on accelerated depreciation.[5] For a mature firm whose acquisitions of new depreciable assets are exactly equal to retirements, annual deductions for normal depreciation will exactly equal outlays for depreciable assets. Gross investment outlays can be wholly financed out of tax-free depreciation allowances. A growing firm,

4. A high discount rate is in many respects similar to a short-pay-off period requirement. A short pay-off period, in fact, implies an infinite rate of discount for returns beyond the pay-off period.

5. This *Journal*, LXVII, 493–519.

however, will be making investment outlays in excess of normal depreciation accruals. Introduction of accelerated depreciation will cause allowable deductions of both the static firm and the growing firm to rise in relation to capital outlays over a period of years. After this transition period, allowable deductions of the static firm will return to equality with investment outlays, but for a steadily growing firm the fraction of investment outlays covered by depreciation allowances will stabilize at a level higher than that obtaining under normal depreciation. For example, a firm growing at a constant annual rate of 5 per cent and employing assets with an average normal life of twenty years will ultimately be able to finance 87 per cent of its investment outlays from tax-free depreciation accruals under five-year straight-line depreciation, as compared with 63 per cent under normal straight-line depreciation. Other things equal, the greater the degree of acceleration of depreciation, the larger will be the fraction of capital outlays which a growing firm can cover from internal funds.[6]

Furthermore, accelerated depreciation will make it easier for firms which are inclined to do so to raise outside capital. Prospective creditors will usually insist that a loan contracted to finance purchase of plant or equipment be repaid in a period shorter than the expected useful life of the facility. Adoption of accelerated depreciation will therefore lessen the danger that the income tax will interfere with repayment of the loan. Equity investors may be willing to pay more for shares if they feel that accelerated depreciation improves the profit outlook.

C. *Factors Affecting Influence of Accelerated Depreciation*

The influence of any given accelerated depreciation scheme depends above all on the attitudes of investors as reflected in their subjective evaluations of time discount and of risk and uncertainty. Other, less individualistic, factors are also significant.

Tax Rate. Among the latter factors probably the most important are the present and expected future rates of tax on net profits. Other things equal, the higher the tax rate the more the investor

6. The fraction of capital outlays of the growing firm which can be covered from depreciation accruals will be an inverse function of the rate of growth and the depreciation period. When accelerated depreciation is introduced, depreciation will immediately rise in relation to investment outlays and will reach a final magnitude after a period of years equal to the average normal life of its depreciable property.

It is assumed that average normal life for tax purposes equals average useful life in actual practice, that is, that assets are not discarded before the end of their normal lives.

gains from accelerated depreciation. Accelerated depreciation is merely a method of liberalizing deductions allowed in computing taxable income, and the advantages of all such schemes are directly related to the tax rate.

Up to this point it has been assumed that the tax rate is constant over the whole life of the asset. As regards any one asset, however, it is clear that acceleration of depreciation allowances will be most advantageous to investors when tax rates are expected to fall in the future and least advantageous when tax rates are expected to rise. In the latter case, the greater amount of future tax savings under normal depreciation accounting will wholly or partly offset the interest advantage and greater certainty of earlier but smaller tax savings associated with accelerated depreciation. Nevertheless, the expectation of rising tax rates does not necessarily diminish the attractiveness of accelerated depreciation to a stable or growing firm, since the firm's annual allowances on its plant and equipment as a whole will always be at least as great under accelerated depreciation as under normal depreciation.

Capacity to Absorb Deductions. Depreciation allowances of any type are valuable to the taxpayer only to the extent that they can be offset against taxable income. An accelerated depreciation plan that produces deductions in excess of current taxable income will be harmful unless the taxpayer can carry back the accounting loss against taxable income of prior years or carry it forward to future years (within the pay-off period). This indicates that the effectiveness of accelerated depreciation will be increased by liberal provisions for averaging of accounting losses and profits. In the absence of such provisions, mandatory accelerated depreciation might actually be a deterrent to investment in time of depression. An excessive rate of acceleration may have this disadvantage at any time. An optional scheme will not suffer from this defect.

As noted above, however, a liberal carryback of losses — as distinguished from a carryforward — may lessen the need for accelerated depreciation, because it will decrease the probability that the income tax will interfere with recovery of capital even with normal depreciation allowances. The existence of a carryback will not reduce other advantages associated with accelerated depreciation, that is, the interest or time-discount gain and the assistance afforded to internal financing. Indeed, in a bad year, the combination of accelerated depreciation and a carryback of losses may result in tax refunds which will raise internal funds to a higher level than would be attained if there were no income tax.

Business Conditions. Like any other scheme which relies mainly on removal of deterrents to investment rather than on the creation of positive incentives, accelerated depreciation is likely to be less effective in time of depression than during prosperity. This characteristic will be accentuated if, in pursuance of a countercyclical policy, tax rates are reduced in time of depression and raised during prosperity.

Capital Structure. Accelerated depreciation will, of course, be most significant for those firms for which it constitutes the largest fraction of total costs. Usually these will be firms in capital-intensive industries and industries in which the ratio of depreciable assets to total investment is high. Accelerated depreciation, therefore, will be more helpful to heavy industry — and to real estate, if buildings are eligible — than to service industries and trade.

D. *Revenue Effects*

Like a reduction in tax rates, adoption of accelerated depreciation will decrease government revenue unless the action stimulates a net increase in activity large enough to offset the reduction in the effective rate of taxation. Many discussions, however, are based on the mistaken belief that accelerated depreciation results in only a postponement of tax revenue with no permanent loss to the government. The error arises from concentrating on a single asset rather than the flow of investment over time. For any one asset, increased deductions in early years are exactly offset by decreased deductions in later years.[7] If tax rates are constant, tax liabilities with respect to income produced by the asset over the period as a whole are not affected. But by the time annual allowances on this first asset decline below the level of normal allowances or disappear, new assets will have become eligible for accelerated depreciation. If both tax rates and the annual volume of investment in eligible assets (for replacement and expansion) remain constant, introduction of accelerated depreciation will cause an initial revenue loss which will never be recovered. Tax collections will gradually return to their prior annual level but, under the assumed conditions, will go no higher and hence will provide no offset to the initial revenue loss. If the volume of eligible investment grows over time — owing to causes other than the allowance of accelerated depreciation — the aggregate revenue loss due to the scheme will also increase.

The revenue effects of accelerated depreciation may to some

---

7. Neglecting the fact that some firms will go out of business or cease to have taxable income after the end of the accelerated depreciation period but before the end of the normal depreciation period.

extent reinforce cyclical fluctuations in business activity. Taxable income and tax liabilities will tend to be lower in years when investment is large and higher in years of depression than they would be under normal depreciation. Under plans other than those confined to an additional allowance in the first year, however, accelerated deductions associated with the investment carried out during prosperity will continue through the beginning of a downturn in activity and may be helpful at that stage, especially if liberal carrybacks of losses are permitted.

E. *Comparison with Reduction of Tax Rates*

Accelerated depreciation is in many respects similar to a selective reduction of tax rates. The tax reduction — or postponement — is offered only to investors in eligible assets and the amount of tax relief depends directly on the amount of such investment. Taxes on income from old capital and from other sources are not reduced.

One way of placing in perspective the possible incentive effects of accelerated depreciation is to calculate what reductions in tax rates would have to be made to improve the expected return from a new investment — or to increase the value of the asset — to the same degree as would adoption of a specified depreciation plan. From the point of view of an investor who accepts the underlying assumptions

TABLE I

TAX RATES WITH NORMAL STRAIGHT-LINE DEPRECIATION EQUIVALENT TO 50 PER CENT TAX RATE WITH FIVE-YEAR AMORTIZATION: TEN-YEAR ASSET, 6 PER CENT INTEREST RATE[1]

| Gross Annual Yield[2] | Tax Rate Equivalent to Five-Year Amortization | | |
| | | Five-Year Pay-Off | |
	Ten-Year Pay-Off	No Carryback	Two-Year Carryback[3]
100	49.2	44.4	46.1–49.3
80	49.0	42.9	45.0–49.1
60	48.6	40.0	42.8–48.6
40	47.6	33.3	37.4–47.3
30	46.4	25.0	29.9–44.9
27.5[4]	45.9	21.4	26.3–43.4
25	45.2	[5]	21.3–35.6
20	42.8	[5]	[5]

[1]Two arrangements are considered "equivalent" if they permit the same present value for an asset of given gross yield. It is assumed that gross yield is the same amount each year, that the tax rate is constant, that all yields and taxes accrue at the end of the year, and that the asset is expected to have no salvage value when retired.
[2]Before depreciation allowances and taxes, as per cent of original cost.
[3]The lower figures are based on the assumption that the asset is retained in use through the sixth and seventh years but produces no income; the higher figures, on the assumption that the asset is discarded at the end of the fifth year with zero salvage value.
[4]Minimum acceptable yield with five-year pay-off, 50 per cent tax rate, five-year amortization, and no loss carryback; at this yield the present value of the asset equals its cost under the assumed conditions.
[5]Investment not acceptable with 50 per cent tax rate and five-year amortization.

regarding the appropriate pay-off period, interest rate, and other relevant factors, these tax reductions and accelerated depreciation are equivalent measures. For another investor, acting on different planning assumptions, accelerated depreciation will be equivalent to a different amount of tax reduction. The state must also regard introduction of accelerated depreciation as a tax reduction, but calculations of the type presented here shed no light on the amount of revenue loss.

It is again convenient to take for illustration an asset with a ten-year normal useful life. Table I shows for different gross yields on such an asset what tax rates in conjunction with normal straight-line depreciation will be equivalent to a 50 per cent tax rate in combination with five-year amortization. Two arrangements are deemed to be equivalent if they permit the same present value for the asset at any given gross yield and on the basis of the assumed discount rate. The table deals with the earnings and value of only one asset, no account being taken of the effect of a change in tax rates on income from other sources or of the possibility of deducting depreciation charges on the asset against income from other sources.

The tax rates entered under the column headed "Ten-year pay-off" reflect solely the interest advantage of early depreciation deductions. It appears that when the gross yield of investment is much above the margin the pure time-discount gain associated with accelerated depreciation could be duplicated by a fairly small reduction in tax rates. These calculations, however, may understate the degree to which liberalized depreciation allowances can ease the financing problem of a growing business, because conventional interest rates do not reflect the full value of internal funds to a growing firm which lacks ready access to external funds.

On the other hand, the opportunity of charging off all or a large fraction of the cost of an asset against income earned in the early years of use may be highly important for an investor who, as an allowance for risk, adopts a subjective pay-off period much shorter than the normal useful life of the asset and who does not take into account the possibility of completing amortization of the asset out of income that may be realized from other sources after the end of the pay-off period. This opportunity can be provided by accelerated depreciation or, in some instances, by a loss carryback. In Table I the significance of accelerated depreciation and of a loss carryback is illustrated by the figures for a five-year pay-off period. In the absence of a loss carryback, accelerated depreciation can be a powerful influence. For

example, on the assumptions underlying Table I, adoption of five-year amortization would be equivalent to halving the tax rate on an asset expected to earn a gross return of 30 per cent a year. If a generous loss carryback is already in effect, however, addition of accelerated depreciation may be comparatively unimportant for many investments. As can be seen from the range of figures presented in the last column of Table I, the investor's attitude is likely to be greatly affected by the particular assumption he makes as to how the carryback will operate. Although it is difficult to generalize on this point, the figures which minimize the net additional contribution of accelerated depreciation in conjunction with a carryback (that is, the higher figures in the last column of Table I) seem to represent the most plausible single assumption.[8] These figures nevertheless overstate to some degree the significance of the carryback because they imply that its operation is absolutely certain and that the investor will take maximum advantage of it. Even on these assumptions, the existence of a carryback covering a large fraction of the pay-off period does not destroy the significance of accelerated depreciation as a means of reducing investment risks. Accelerated depreciation may still be equivalent to a large tax reduction for an asset whose expected gross yield is only a little above the margin (see the figures for assets yielding 25–30 per cent).

One of the most interesting features of accelerated depreciation, regardless of whether combined with a loss carryback, is the fact that it is relatively more important for low-yield investments than for high-yield assets. This is true because the additional depreciation deductions represent a much larger proportion of the total return from a low-yield asset than from a high-yield investment. The most promising means of stimulating investment is to bring within the range of favorable consideration projects which under given circumstances lie just outside the range. There is little advantage in increasing the attractiveness of projects which already promise a high net yield. Hence the capacity of accelerated depreciation to have the same influence as a large tax cut for marginal projects and a small cut for highly lucrative projects is a great advantage.

A few words should be added about the significance of the fact that the illustrative computations summarized in Table I relate to

8. If the investor retains the asset in a standby capacity after the end of the pay-off period and thereby foregoes the opportunity of immediately charging off the unrecovered balance of cost, he presumably anticipates that a situation may arise in the future in which the asset will earn enough to compensate for the smaller loss carryback and smaller tax refund.

one asset viewed more or less in isolation. As already explained, this approach may exaggerate the contribution of accelerated depreciation as a corrective for a discrepancy between the lengths of the pay-off period and the normal depreciation period, because it does not allow for the possibility of delayed amortization against other income. But the approach is less favorable to accelerated depreciation in that it ignores an important comparative advantage of the measure due to the fact that it can easily be confined to new investment. Since it would not be practicable to restrict a cut in tax rates to income from new investment, a large part of the benefits would go to owners of existing capital goods. This is one reason for believing that sacrifice of a given amount of revenue will be more effective in stimulating investment if it is in the form of acceleration of depreciation allowances than if in the form of a reduction in income tax rates.

F. *Accelerated Depreciation as an Alternative to Other Tax Concessions in Underdeveloped Countries*

Many underdeveloped countries have adopted or considered plans to encourage saving and investment by means of income tax concessions. One method is to offer a temporary tax exemption to new firms or to profits earned on new investments. Considerable difficulty is encountered, however, in identifying a genuinely new business and in allocating the profits of an existing firm between old and new capital. Furthermore, a general exemption of this type is always subject to the objection that it grants disproportionate benefits to firms which enjoy a high rate of profit. Some countries have attempted to avoid this result by confining the exemption to profits not in excess of a stated rate of return on invested capital, but the difficulties of applying such a standard are notorious. Even if the exemption can be successfully limited to a maximum rate of return on invested capital, it will often bear no direct relation to the needs of the firm for new capital.

Another approach is to exempt savings from the income tax or to apply a lower rate to this portion of income. Many schemes designed to carry out this purpose, however, are open to the objection that they grant concessions to hoarding and to savings used to acquire land, buildings, and other types of investment which the government does not wish to encourage. In fact, it is possible to prevent misdirected favors and abuses only by obtaining complete balance sheets showing the assets and liabilities of each taxpayer at the beginning and end of the year. Countries with the most advanced revenue

administrations have not found it feasible to require such information from individuals on a wide scale, although it would be of undoubted value in assessing even an undifferentiated income tax.

Accelerated depreciation appears to be a very promising alternative to both types of schemes. In view of the fact that interest rates are ordinarily high and capital scarce in underdeveloped countries, any measure that speeds the amortization of investment projects and permits greater reliance on internal financing is highly advantageous. Furthermore, as already explained, accelerated depreciation represents a relatively less significant concession for highly profitable firms and for stagnant or declining firms than for marginally profitable or rapidly growing firms. (All types of firms will benefit but the amount of benefit will be related to the amount of investment, rather than to the size of profits as under a tax exemption.) This differentiation is automatically achieved by a technique which is much simpler than any plan for carefully restricted temporary tax exemption. The availability of tax relief under accelerated depreciation depends on the behavior of the individual firm, and the amount of benefits is directly related to the rate of expansion as measured by acquisition of new depreciable assets. A rapidly growing firm whose investments yield a moderate rate of return may enjoy an extended period of virtual exemption from the income tax, whereas a static or declining firm will pay, after the initial transition period, the same amount of tax as it would with normal depreciation. Only a limited number of special administrative problems are associated with accelerated depreciation.[9] Any country that can successfully administer a tax on net income from business can apply an accelerated depreciation plan. Accelerated depreciation is a less complete alternative to tax exemption for individual savings, but it has an important advantage in that the benefits are automatically restricted to those who invest in approved productive facilities. Accelerated depreciation can be granted to types of investment which the government wishes to promote and withheld from other types of investment.

Of course, accelerated depreciation will be a significant conces-

---

9. Chief of these special problems is to prevent double allowances with respect to property which, after being fully amortized, is transferred to another firm in order to obtain a second series of depreciation deductions. The simplest way of dealing with this type of abuse would be to tax at regular rates any gain that the seller realizes on property which has been eligible for accelerated depreciation, the gain to be measured by the difference between the depreciated (book) value of the asset and the sales proceeds. This rule would eliminate opportunities for multiple tax benefits without interfering with normal transfers.

sion only in countries where effective income tax rates are high enough to influence savings and investment. It will assist only those firms which are able to realize a profit fairly soon after making a new investment. The same limitations hold for all plans for a limited exemption from the income tax.

G. *Comparison with Reduction of Interest Rates*

Adoption of accelerated depreciation may be compared with a reduction of interest rates as well as with a cut in tax rates on business profits. A reduction of interest rates, like the tax revisions, increases the net yield and present value of capital assets and hence may act as a stimulus to investment. Tax policy and monetary policy are in this respect alternative means of promoting investment. They are, of course, not perfect substitutes. A change in interest rates is likely to be general in scope and to influence directly the market values of assets such as bonds, residential housing, and land as well as business plant and equipment. Furthermore, the interest rate may have a more direct effect on the community's willingness to save and to lend.

The similarity between introduction of accelerated depreciation and a reduction in interest rates will be most apparent to investors who adopt a sophisticated planning procedure involving imputation of interest costs to their own capital as well as to borrowed funds. Such a procedure will also logically imply that in valuing investments their expected future yield is subjected to time discount at the imputed interest rate. In practice many investors probably make no systematic allowance for interest on their own capital or for time discount on deferred income, and those who do make an allowance may apply a conventional interest rate which responds only slowly to changes in market rates. Investments undertaken by these investors will be to some extent insulated against the effects of monetary policy. The investors who do not allow for imputed interest on their own capital can hardly be expected to make a formal calculation of the time-discount gain associated with accelerated depreciation. They may nevertheless recognize an advantage in early recovery of capital and they may react favorably to the improved opportunities for internal financing offered to growing firms.

The time-discount gain is the feature of accelerated depreciation which is most directly comparable to a reduction in market rates of interest. In some instances it may be possible to compensate for risk and uncertainty by an adjustment of interest rates, but in general the

possibilities are limited. When the investor applies an over-all discount rate covering both risk and interest, the interest component will usually be small relative to the risk component; hence, the over-all rate will typically be rather insensitive to changes in market rates of interest. When the allowance for risk and uncertainty takes the form of a short pay-off period, additional depreciation deductions brought within the pay-off period as a result of an accelerated depreciation scheme are likely to outweigh the effects of feasible changes in interest rates.

TABLE II

Discount Rates Which Make Normal Depreciation Equivalent to Five-Year Amortization and 6 Per Cent Discount Rate: Ten-Year Asset, 50 Per Cent Tax Rate[1]

Gross Annual Yield[2]	Discount Rate
100	5.73
80	5.67
60	5.57
40	5.41
30	5.26
20	5.00

[1] A ten-year pay-off period is assumed. Two arrangements are considered "equivalent" if they permit the same present value for an asset of given gross yield. It is assumed that gross yield is the same amount each year, that all yields and taxes accrue at the end of the year, and that the asset is expected to have no salvage value when retired.
[2] Before depreciation allowances and taxes, as per cent of original cost.

Some calculations illustrating the comparative significance of reduction of interest rates and the time-discount gain associated with accelerated depreciation are presented in Table II. This table shows the rates of discount which, when applied to the expected future yield of assets subject to normal depreciation, will result in the same present values for the assets as would obtain with accelerated depreciation and a 6 per cent discount rate. The tax rate is held constant in these calculations. The calculations are based on the assumptions that investors impute the same rate of interest on their own capital and on borrowed capital and that they subject expected future income to time discount at this rate. Since the pay-off period is assumed to be equal to the normal depreciation period, the figures isolate the time-discount feature of accelerated depreciation. (The figures in Table II therefore may be compared with those appearing in the "Ten-year pay-off" column of Table I but are not comparable with the tax rates shown in the "Five-year pay-off" columns of Table I.)

Table II again brings out the point that accelerated depreciation

is more significant for assets with moderate yields than for those with very high yields. Although the absolute amount of the time-discount gain depends on the degree of acceleration of deductions and the interest rate rather than on the expected gross yield of the asset, the gain represents a larger fraction of the total return on a moderate yield asset. Hence a greater reduction of interest rates would be required to duplicate the influence on the present value of an asset of moderate yield. This appears to be an advantage of accelerated depreciation as compared with a reduction in interest rates. The fact that accelerated depreciation does not extend to other types of investment such as inventories, land, and bonds which would be affected by a general reduction in interest rates may also be an advantage in many circumstances. Another significant feature of accelerated depreciation is that, like other tax reductions, it can have the same effect on the value of selected capital goods as a cut in interest rates but without reducing the reward for saving and lending.

II. Comparison of Accelerated-Depreciation Methods

In order to compare alternative methods of granting depreciation allowances for tax purposes, six schemes will be briefly examined:

1. Normal straight-line depreciation.
2. Five-year amortization with allowances distributed on a straight-line basis.
3. Declining-balance depreciation at an annual rate twice the normal straight-line rate, with the taxpayer allowed the option of shifting to the straight-line method for the unrecovered balance of cost at any time.
4. Declining-balance depreciation at an annual rate equal to five-fourths of that rate which would write down the value of the asset to approximately 10 per cent of original cost at the end of its normal useful life.[1]
5. An initial allowance of 20 per cent plus the declining-balance annual rate allowed under method 4.
6. An investment allowance of 20 per cent plus the declining-balance annual rate allowed under method 4.

As previously explained, the straight-line methods (numbers 1 and 2) provide an equal amount of depreciation allowance each year. The normal procedure (method 1) is to compute this allowance by dividing the original cost of the asset, less any estimated salvage value, by the number of years of expected useful life. Under method 2 the annual allowance is computed by dividing by five even though

1. The annual rate $= 1.25 \left[1 - \left(\dfrac{10}{100} \right)^{\frac{1}{n}} \right]$, where $n =$ the number of years of normal useful life.

the normal useful life of the asset exceeds five years. The declining-balance methods (numbers 3 and 4) permit the taxpayer to deduct each year a constant percentage of the unrecovered balance of his capital outlay; since the constant percentage is applied to a declining balance, the amount of the deduction decreases each year. The only difference between methods 5 and 6 is that the "initial allowance" granted by method 5 is deducted from the written-down value of the asset, which serves as the basis for subsequent annual allowances, whereas the "investment allowance" granted by method 6 is not so deducted. This means that the initial allowance is a pure acceleration of normal deductions and that the investment allowance is a net addition to normal deductions.

A. *Use of the Methods*

Straight-line depreciation (method 1) is the procedure that has been normally followed in the United States in the past, although other methods have been employed to a limited extent. Five-year amortization (method 2) was allowed in the United States for all or part of the cost of certified emergency facilities during World War II, and it was reintroduced after the outbreak of the Korean War. Method 3 — a combination of declining-balance and straight-line methods — was authorized by the recent U.S. tax revision act (Internal Revenue Code of 1954). This alternative method is limited to property new in use and never before subject to depreciation allowances, and to be eligible an asset must have a useful life of three or more years. As compared with past normal procedures, it involves a considerable acceleration of depreciation allowances. Both the normal straight-line method and five-year amortization for certified emergency facilities will continue in force.[2]

The version of the declining-balance method which is here labeled "method 4" is the normal procedure in the United Kingdom for machinery and equipment, but not for buildings.[3] The declining-balance method is also employed by other members of the British

---

2. Taxpayers will also be allowed to adopt any other "consistent" method which will result in deductions not in excess of those provided by the approved version of the declining-balance method during the first two-thirds of the service life of the asset. One method specifically authorized is the "sum-of-the-years' digits" method, which produces a pattern of diminishing deductions but, unlike the declining-balance method, allows full recovery of original cost by the end of the asset's normal life. See *Internal Revenue Code of 1954*, sec. 167, and Report of Senate Finance Committee on H.R. 8300 (Report 1622, 83d Congress, 2d session), pp. 27–28.

3. Board of Inland Revenue, *Income Tax Wear and Tear Allowances for Machinery or Plant: List of Percentage Rates* (London: H.M. Stationery Office, 1950), pp. 3–4.

Commonwealth, but it appears that in several of these countries the typical annual rates are lower than in the United Kingdom. In comparing the British and American methods, it should also be borne in mind that estimates of normal useful lives of particular types of assets often differ in the United States and the United Kingdom. The initial-allowance approach (method 5) became effective in the United Kingdom in 1946, was suspended in 1952, and was reintroduced in 1953. The amount of the allowance for machinery and equipment was originally 20 per cent; it was raised to 40 per cent in 1949 but was restored to 20 per cent in 1953. The initial allowance for industrial buildings has been 10 per cent throughout the period during which it has been in force.[4] The investment allowance was introduced in the 1954–55 budget. Although the allowance is at the same rate as the previous initial allowance, for buildings as well as machinery and equipment, the two schemes, as already explained, are fundamentally different. Under the current legislation investors in mining works may choose either a 20 per cent investment allowance or a 40 per cent initial allowance. The initial allowance is continued for secondhand machinery and for "ordinary" motor cars.

B. *Annual Charges and Their Present Value*

Application of the six methods may be illustrated by computing annual depreciation charges for an asset with a ten-year normal useful life. The normal straight-line procedure (method 1) will produce annual charges of 10 per cent of original cost for ten years.[5] Five-year amortization on a straight-line basis (method 2) will allow annual charges of 20 per cent of original cost for five years.[6] The four versions of the declining balance method (methods 3, 4, 5, and 6) will result in annual and cumulative charges as shown in Table III. This table reveals two characteristics of the normal declining-balance method: (1) the concentration of charges in the early years of life of the asset and (2) the fact that a part of the original cost of the asset is not written off by the end of its normal useful life.[7] The latter

4. The 1953 legislation provided a 40 per cent initial allowance for new mining works, including pit shafts, oil wells, etc., but excluding machinery and equipment.

5. Provided the asset is expected to have no salvage value at the end of its useful life.

6. See preceding note.

7. If the unrecovered balance of cost at the end of the normal useful life is approximately equal to the expected salvage value, its existence is not a disadvantage of the declining-balance method as compared with the straight-line method because under the latter method annual deductions are limited to the excess of original cost over estimated salvage value.

feature is avoided under the hybrid method recently authorized in the United States by allowing the taxpayer the option of shifting to the straight-line method at any time for the unrecovered balance of cost. In the illustrative case, this option becomes advantageous at

TABLE III

ALTERNATIVE DECLINING-BALANCE DEPRECIATION METHODS: ANNUAL AND CUMULATIVE CHARGES AS PER CENT OF ORIGINAL COST OF ASSET WITH TEN-YEAR NORMAL LIFE

Year	Normal 25 Per Cent Rate		Hybrid[1]		With 20 Per Cent Initial Allowance[2]		With 20 Per Cent Investment Allowance[2]	
	Annual	Cumulative	Annual	Cumulative	Annual	Cumulative	Annual	Cumulative
1	25.00	25.00	20.00	20.00	45.00	45.00	45.00	45.00
2	18.75	43.75	16.00	36.00	13.75	58.75	18.75	63.75
3	14.06	57.81	12.80	48.80	10.31	69.06	14.06	77.81
4	10.55	68.36	10.24	59.04	7.74	76.80	10.55	88.36
5	7.91	76.27	8.19	67.23	5.80	82.60	7.91	96.27
6	5.93	82.20	6.55	73.79	4.35	86.95	5.93	102.20
7	4.45	86.65	6.55	80.34	3.26	90.21	4.45	106.65
8	3.34	89.99	6.55	86.90	2.45	92.66	3.34	109.99
9	2.50	92.49	6.55	93.45	1.84	94.50	2.50	112.49
10	1.88	94.37	6.55	100.00	1.38	95.88	1.88	114.37
11	1.41	95.78	....	....	1.03	96.91	1.41	115.78
12	1.06	96.84	....	....	.77	97.68	1.06	116.84
13	.79	97.63	....	....	.58	98.26	.79	117.63
14	.59	98.22	....	....	.44	98.70	.59	118.22
15	.44	98.66	....	....	.32	99.02	.44	118.66
16	.34	99.00	....	....	.24	99.26	.34	119.00
17	.25	99.25	....	....	.18	99.44	.25	119.25
18	.19	99.44	....	....	.14	99.58	.19	119.44
19	.14	99.58	....	....	.10	99.68	.14	119.58
20	.10	99.68	....	....	.08	99.76	.10	119.68

[1]Declining-balance method at 20 per cent annual rate with option to shift to straight-line method for unrecovered cost.
[2]Declining-balance depreciation deductions at 25 per cent annual rate.

the end of the sixth year. The investment-allowance device (method 6) does not avoid the "tail" which is characteristic of the declining-balance method, but deductions exceed original cost before the end of the asset's normal life.

Under both the straight-line method and the ordinary declining-balance method, it is customary to allow a final deduction equal to the unrecovered balance of the cost of the asset or group of assets when the asset or the last unit of the group is retired. This final adjustment is equal to the written-down value minus any salvage

value realized on retirement. (Presumably under the investment-allowance plan — method 6 — the final adjustment would be the amount necessary to permit recovery of 120 per cent of original cost.) If, however, the asset is retained in use after the end of its normal life, annual charges of the type shown in Table III for years 11–20 will be allowed.

If we abstract from changes in tax rates and assume that the taxpayer will always have enough taxable income to absorb the annual charges, calculations of the type presented in Table IV shed con-

TABLE IV

TOTAL CHARGES AND PRESENT VALUE OF CHARGES FOR STATED PERIODS UNDER ALTERNATIVE DEPRECIATION METHODS: TEN-YEAR ASSET, 6% INTEREST RATE

(Charges as per cent of original cost)

| Depreciation Method | Total Charges[1] | | | Present Value of Charges[4] | | |
	1st 5 Years	1st 10 Years [2]	[3]	1st 5 Years	1st 10 Years [2]	[3]
1. Normal straight line	50.0	100.0	100.0	42.1	73.6	73.6
2. Five-year amortization	100.0	100.0	100.0	84.2	84.2	84.2
3. Declining balance, 20 per cent rate, with option to shift to straight-line method[5]	67.2	100.0	100.0	58.1	78.7	78.7
4. Declining balance, 25 per cent rate	76.3	94.4	100.0	66.3	78.1	81.2
5. Initial allowance of 20 per cent plus declining-balance depreciation at 25 per cent rate	82.6	95.9	100.0	73.8	82.4	84.8
6. Investment allowance of 20 per cent plus declining-balance depreciation at 25 per cent rate . . .	96.3	114.4	120.0	85.2	97.0	100.1

[1]Including final allowance, if any.
[2]On assumption that asset continues in use.
[3]On assumption that asset is retired at end of period with zero salvage value.
[4]Including final allowance, if any, and assuming that annual charges accrue at end of each year.
[5]Shift to straight-line method (for unrecovered balance) assumed to occur in seventh year.

siderable light on the possible influence of the different depreciation methods. This table shows the total allowable depreciation charges and their present value (at the time when the investment is made) for the first half of the life of an asset with a normal durability of ten years and for the whole normal life. The present value of cumulative charges at the end of each year is shown graphically in the accompanying chart. The figures on charges can be translated into tax savings by multiplying them by the tax rate.[8] With a 50 per cent tax rate,

8. On the assumption that both charges and taxes accrue at the end of each year. The calculation is, of course, simplest with a proportional tax rate, which is assumed in all of the illustrations in this paper.

for example, the present value of the tax saving is equal to one-half of the figures shown in the last three columns of Table IV. Other things the same, a difference in the present value of the tax savings attributable to the charges under two depreciation methods will be reflected in an equal difference between the present value of an asset subject to one of the methods and the present value of an identical asset subject to the other method.

PRESENT VALUE OF CUMULATIVE CHARGES UNDER ALTERNATIVE DEPRECIATION METHODS:
10 YEAR ASSET, 6% INTEREST RATE
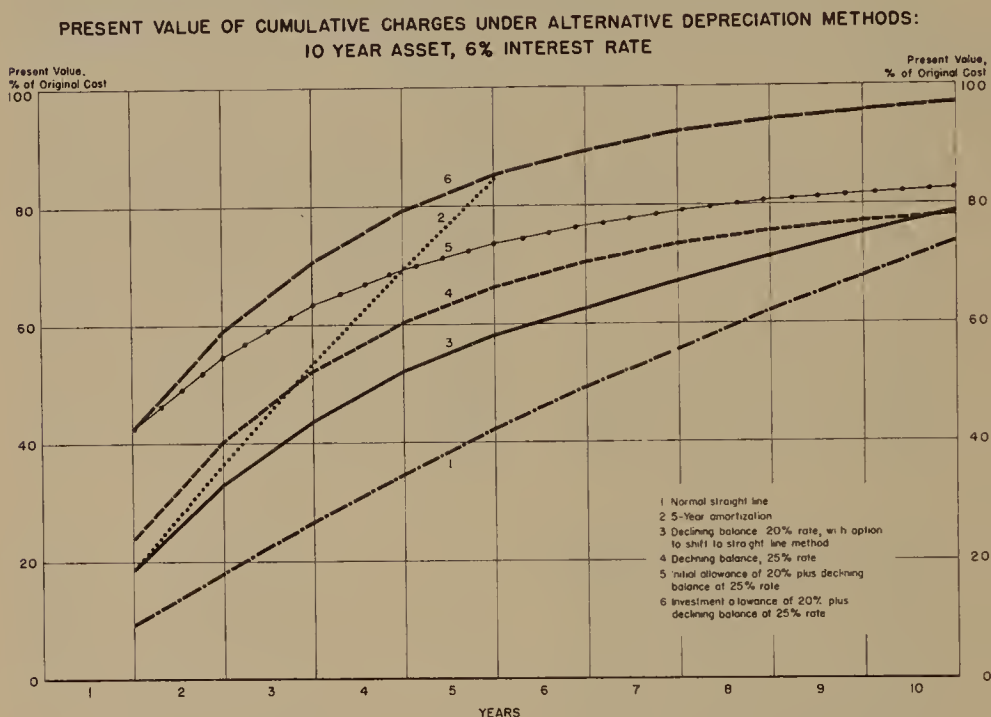
Of the methods here compared only five-year amortization (method 2) and the investment allowance (method 6) permit the investor to recover free of tax all or substantially all of the cost of a ten-year asset in the first half of its useful life. To the extent that investors insist on pay-off periods much shorter than the normal useful life of an asset, this feature is a distinct advantage of these two methods from the taxpayer's point of view. During the second half of the asset's normal life the liberality of the investment-allowance plan becomes evident, but it should be remembered that this is due entirely to the fact that deductions exceed original cost which, of course, implies a greater revenue sacrifice for the government than is involved in the plans limited to pure acceleration of allowances.

An interesting point is that, on the assumption of a 6 per cent discount rate, the hybrid method (number 3) recently adopted in the

United States is somewhat less liberal than the normal British method (number 4). It does not necessarily follow, however, that in practice British depreciation allowances are more generous than American allowances. Such an inference would be valid only if the British authorities ordinarily assumed that useful lives of different types of assets were no longer than the lives estimated for comparable items by the American authorities. It is difficult to ascertain the facts on this point, but it appears that, with some important exceptions, the usual British practice is to assume somewhat longer useful lives for machinery and appreciably longer lives for buildings.[9] Nevertheless, it appears that from the point of view of investors who apply a time discount to expected income, setting a somewhat higher annual rate under the normal declining-balance method may be a fairly effective substitute for the option to switch to the straight-line method, as provided by the recent U.S. law. This is true because the time-discount factor tends to offset the fact that, under the declining-balance method, part of original cost remains unrecovered so long as the asset continues in use. Investors who do not systematically discount future income may, however, attach more importance to full recovery of costs than to early recovery and hence prefer the U.S. method.

Five-year amortization and a uniform initial allowance or investment allowance all favor long-lived assets as compared with short-lived assets. This characteristic is most evident in the case of five-year amortization, which grants no benefit to investment in assets with a normal life of five years or less and a large benefit to investment in more durable property. The initial allowances and investment allowances, if equal to the same percentage of original cost for assets of different durability, also increase the present value of long-lived items considerably more than the present value of short-lived assets. But for very long-lived assets initial allowances or investment allowances on the scale granted in the United Kingdom are much less generous than five-year amortization.[1]

9. The generalization regarding machinery is based on an examination of lists of typical rates for various industries and items presented in U.S. Bureau of Internal Revenue, Bulletin "F," *Income Tax: Depreciation and Obsolescence, Estimated Useful Lives and Depreciation Rates* (Washington: Government Printing Office, 1942) and U.K. Board of Inland Revenue, *Income Tax Wear and Tear Allowances for Machinery or Plant: List of Percentage Rates* (London: H. M. Stationery Office, 1950). Since, however, the classifications employed in the two official bulletins are different, an exact comparison cannot be made.

1. Domar (this *Journal*, LXVII, 505–6) considers the American five-year amortization scheme unduly generous as compared with the British initial-allowance approach, but, of course, this is a matter of judgment on which it would be difficult to obtain general agreement.

The favoritism to long-lived investments may be considered appropriate on the grounds that such investments are more risky and hence more likely to be discouraged by taxation than investment in shorter-lived property. It may be, however, that the differential impact is unintended. If desired, the discrimination between long- and short-lived assets could be substantially reduced by relating the additional deduction to the normal deduction, which reflects the durability of the asset. This could be easily accomplished by authorizing application of a uniform multiplier to the normal depreciation rate or by stating the initial allowance or investment allowance as a multiple of the first-year's normal deduction.

C. *Appraisal*

In many ways the neatest and simplest plan for accelerating depreciation allowances is merely to increase the rates permitted under normal procedures, whether the straight-line or declining-balance method. Other approaches, however, seem to have had more appeal, possibly because they are more easily recognized as special concessions or because they are believed to yield closer approximations of the time distribution of loss in real economic value of depreciable assets.

The advantages of accelerated depreciation can be largely achieved under all of the methods considered in this paper, and there seems to be no strong reason for preferring one method over other available methods. A fair degree of flexibility can be introduced in all of the methods. For example, five-year amortization has been applied in the United States to varying fractions of the cost of defense-supporting projects rather than to the whole cost of all approved projects, and different initial allowances have been assigned to different industries in the United Kingdom. Any accelerated depreciation scheme, however, will be ineffective to the extent that the taxpayer does not have enough income to absorb the extra deductions allowed during the early years of life of an asset. This limitation seems likely to be especially significant in the case of initial allowances and investment allowances. In order to accomplish a significant degree of acceleration under these methods, it is necessary that the allowance be large relative to the yield that can usually be expected in the year in which an investment is made. The taxpayer, of course, may be able to charge off the allowance against income from other sources,

but large established firms are more likely to be able to do so than small new firms. These considerations suggest that the optimum plan may not be that which grants the largest immediate deductions.

III. CONCLUSION

Acceleration of depreciation allowances seems to be an effective means of reducing the extent to which a tax on income or profits discourages investment. This judgment is based on the opinion that a large part of the deterrent influence of the tax is due to the fact that, with normal depreciation allowances, it delays and in some instances prevents full recovery of capital out of earnings from a new asset. The speedier recovery of capital made possible by accelerated depreciation offers an interest or time-discount gain to investors and permits growing firms to finance more of their capital requirements from earnings. It also lessens risk and uncertainty.

Introduction of accelerated depreciation stimulates investment by lowering a tax obstacle rather than by creating entirely new incentives. The potential influence of accelerated depreciation depends, therefore, on how serious are the obstacles presented by the income tax with normal depreciation methods. In part this is a matter of the tax rate and other statutory provisions. High tax rates obviously influence investment returns more than low rates do, and the advantages of accelerated depreciation are greater under high tax rates. When liberal provisions for carryback of operating losses reduce the risk that the income tax will prevent capital recovery, the net contribution that accelerated depreciation can make is smaller than when the income tax lacks these refinements. The attitude of investors is also important. Accelerated depreciation will be most important to those investors who apply a fairly heavy but not excessive discount for interest and risk or who adopt a pay-off period considerably shorter than the normal useful life of assets but still long enough to permit recovery of a substantial fraction of investment outlay during the pay-off period by means of accelerated allowances.

Investment can also be stimulated by cutting tax rates or by bringing down interest rates. Accelerated depreciation has the advantage of being more selective in that the benefits are restricted to those who acquire new depreciable assets. In proportion to revenue loss it will offer a greater stimulus to investment than a general reduction in income tax rates, and as compared with lower interest rates it will be less likely to stimulate types of investment that the government

may not wish to promote. For underdeveloped countries, accelerated depreciation appears to be a promising alternative to tax-exemption schemes intended to encourage investment.

Since taxation is only one of many factors that may hold back investment, accelerated depreciation alone will often not be enough to assure a desired level of investment. But accelerated depreciation can be a highly useful part of a program to stimulate investment for purposes of income stabilization or economic growth.

RICHARD GOODE.

INTERNATIONAL MONETARY FUND

LUNDBERG ON BUSINESS CYCLES
AND PUBLIC POLICY

By SVEND LAURSEN

Professor Erik Lundberg, author of the celebrated *Studies in the Theory of Economic Expansion*, needs no introduction to an American audience. In an important book published recently[1] he pursues further his work on the business cycle. But while the previous book was concerned primarily with formal model building, the main emphasis in the present volume is on questions of economic policy. The chief strength lies in a masterful blend of analysis and description, both historical and statistical, rather than in new insights or theoretical innovations.

In 1948 Lundberg was requested by a private educational organization to undertake a systematic study of the Swedish experience with economic controls. In the process of preparing the report the scope was widened rather extensively in two directions: first, instead of dealing exclusively with direct controls the final report encompasses an analysis and comparison of a variety of policy instruments; and second, the illustrative material used in the discussion of alternative policies developed into a full-bodied treatment of Swedish business cycles in their international setting, with short excursions into the field of cycle theory. While the historical part (Part I in the book), though illuminating, does not give rise to many controversies, Part II on policy is of a more controversial and polemical character. The lively discussion of the book, which is now taking place in Sweden, centers almost entirely on this portion.[2] In the following comments, I shall also limit myself largely to policy questions.

I

Part I, which presents the basic material for Lundberg's analysis and policy conclusions, draws heavily on the work of the Swedish

---

1. Erik Lundberg, *Konjunkturer och ekonomisk politik* (Stockholm: Konjunkturinstitutet — Studieförbundet Näringsliv och Samhälle, 1953).
2. Almost an entire issue of *Ekonomisk Tidskrift* (Vol. LV, Dec. 1953) was devoted to a series of reviews of the book with a reply by Lundberg to his critics.

Business Cycle Institute of which he is the head. A fairly detailed picture is given of trends and cycles in the Swedish economy since World War I. But parallel with this description runs another one dealing with the fluctuations in major areas of the world with special attention to business cycles in the United States. The reason for these parallel descriptions can be found in Lundberg's thesis that Sweden does not have autonomous cycles. While the dominating influence of external factors has synchronized Swedish and international business cycles, the special structure of the economy and the policies pursued by the government has given these "imported" cycles some characteristics of their own. In particular, during the interwar period the amplitude of the ups and downs was smaller than in the United States and many other countries. Especially, the rising trend in the construction industry contributed much to the stability of total domestic investment.

While Lundberg uses the multiplier in analyzing the impact of changing exports and investment, he warns the reader against a mechanical application of this tool (a similar remark also applies to his use of the acceleration principle). The various propensities are apt to change during the cycle, in part because of shifts in income distribution, and fluctuating proceeds in the export industries will not only affect investment in these industries directly, but also investment in other sectors of the economy, both through income effects and through a change in the liquidity position caused by variations in the foreign balance. As a result of these and other refinements the analysis achieves a high degree of realism.

Perhaps of special interest to a foreign reader is Lundberg's discussion of the thirties as compared with the twenties. The thirties was *not* a period of full employment in Sweden and in this respect the postwar performance has been much better as is indicated by the following table (p. 132):

Years	Average Unemployment as Percentage of Organized Workers
1923–30	11%
1930–33	19
1933–37	16
1937–39	10
1945–50	3

As Lundberg points out the decade of the twenties is often described as a conservative period with restrictive monetary and financial policies and with stagnation in the field of social legislation. The thirties, on the other hand, is looked upon as a progressive period when expansionary policies were pursued and new advances made on

the front of social legislation. As a corrective to this picture he emphasizes that among the merits of the former decade were the return to the gold standard, stable money values, healthy competition which eliminated high-cost producers, and finally the rapid expansion of foreign trade. On the debit side of the thirties, on the other hand, are the abandonment of the gold standard, the fall in the value of money in the second half of the decade, relative stagnation of exports, and a higher degree of self-sufficiency.

What caused, then, the limited upswing in the thirties? Lundberg demonstrates that the importance of deficit spending was negligible and that a major factor in the development was the 1931 depreciation. Although this latter led to only a moderate expansion of exports, he says, its main significance was an increased protection of domestic industry which had previously been subject to hard competition from abroad. Combined with various agricultural measures, the domestic economy became a sheltered one. But further, the annual surplus on international accounts provided a margin for the domestic upswing. And finally, this surplus also increased the liquidity position of business since the foreign exchange earnings were largely taken over by the *Riksbank*. The level of interest rates was reduced with a stimulating effect on investment, particularly in plant and residential housing.

Let me quote Lundberg's appraisal of this period:

"The fine thing about Swedish economic policy during the expansion of the thirties — and for this the authorities have not been praised sufficiently — is primarily to be found in the fact that mistakes made by other countries were avoided. ... It was perhaps fortunate that the fruitful economic-political discussion in Sweden during the thirties was not too greatly interfered with by practical experiments. This was perhaps a good thing both for the development of economic theory and for the development of the actual business cycle. At any rate the Swedish discussion and the theoretical contributions in this field achieved international fame. And mistakenly the outside world often correlated the theoretical innovations by Swedish economists with the rapid and relatively stable expansion in Sweden."[3]

The final chapter in Part I deals with the problems of the postwar period. It is demonstrated that while such aggregates as GNP and industrial production have shown steady growth, and full employment has been achieved and maintained, fluctuations have nevertheless taken place within the margin of full employment. Such indicators as the number of applicants per 100 job vacancies, or the ratio between building permits granted and total applications, have displayed cyclical variations. The major source of instability, however,

3. *Op. cit.*, p. 257.

has been foreign price impulses. In this connection Lundberg makes some interesting comments about "the propensity to inflate" of the Swedish economy. With a starting point in the post-Korean inflation, he constructs a model assuming an equal rise of export and import prices by 50 per cent. On the basis of the importance of imports, a rise of this magnitude can be expected to lead to an overall price increase of approximately 10 per cent when the foreign impulse has spread vertically through the price system. There is, however, a second route through which the impact is felt: similar products produced and sold in the domestic market will rise in sympathy with internationally traded goods. This he calls the horizontal price effect. Moreover, both the vertical and horizontal price increases lead to rising profits and since by assumption the terms of trade have not deteriorated a margin exists for wage increases. That is to say, in principle a certain wage increase is possible and necessary in order to maintain the previous distribution between wages and profits without causing further inflation.

The inflationary process will, of course, be determined in part by the monetary policy pursued by the banking system. If credit is not extended too freely, certain forces will automatically develop which tend to act as a brake on the inflation. High marginal tax rates and increased saving out of higher profits from inflation, according to Lundberg, will absorb some of the additional purchasing power, and tension may thus develop between demand and the inflated cost level.

II

In his discussion of public policy Lundberg's point of departure is a classification of goals according to their degree of ambition. At one end of the scale is the classical gold standard which makes few demands on the government because of the supposed automaticity of the system; at the other extreme is a policy designed to achieve not only full utilization of all resources but also their optimal utilizations. In between are a whole series of alternatives varying according to aspirations (e.g., a stable price level, reduced cyclical fluctuations, full employment, etc.). It appears that his own preferences are in favor of relatively modest aims in economic policy. He rejects in summary fashion modern welfare economics as a guide to policies, but his reluctance to accept more ambitious goals can first of all be traced back to his view that most of the available tools are highly imperfect. He also admits, however, that in the final analysis his own views on policies, as well as those of others, must of necessity

contain ideological elements. In these matters it is not possible to remain within the realm of pure science.

In connection with the classification of alternative goals, Lundberg makes the well-known point that the goals may be in conflict with one another. While this is perfectly true and is important to emphasize from a conceptual point of view, his presentation runs the risk of being misleading. It is true that in the Swedish discussion in particular, the supposed conflict between stable prices and full employment has been very much in the foreground. It is not correct, however, from this to draw the conclusion that actual policy decisions and choices in the postwar period have been guided by considerations of such conflicts. It would be more accurate to state that in Sweden, as elsewhere, the level of effective demand has been sufficient, or more than sufficient, to bring about full employment without conscious policy measures. If this is correct it also follows that a full employment policy, which is the proclaimed aim of many governments, has not as yet been subjected to a real test. Without going into details, all the major postwar decisions in Sweden — such as the lowering of the interest rate in 1945, open market operations in 1949, and the abolishment of the sales tax — have been made with little regard to the level of employment and also, it seems, with little awareness of the Lundbergian conflicts. But to point to these facts does not, of course, detract from Lundberg's observation that goals are not always reconcilable.

III

Throughout the book it is repeatedly emphasized that both the working of the price mechanism and the efficiency of any particular policy measure depend significantly upon the total economic situation. For example, in cases of major economic disturbances and imbalances, less reliance can be placed upon the price mechanism as an equilibrator than when deviations from equilibrium are of a small order of magnitude. Accordingly, Lundberg feels that one of the chief preoccupations of public policy should be with the elimination of such imbalances, in order to create conditions favorable to the functioning of the price system. While this view appears eminently sound he seems, on occasion, either to forget it or at least to play it down. In dealing with problems of the balance of payments he thus advocates greater flexibility for exchange rates, but leaves out any serious consideration of the very real possibility that such arrangements may not always work because of destabilizing speculation caused by an initial imbalance. But further, in his severe criticism

of direct controls it sometimes appears that he ignores his own view that in certain situations no other avenues may be open. And finally, in criticizing such controls he leaves himself open to the criticism that an insufficient distinction is made between, say, the detrimental effects of inflation and the negative aspects of a control system. Conceptually the two should not be identified as Lundberg tends to do.

My second general comment refers to the relationship between ends and means. It is inspired by an important recent essay by Tinbergen,[4] who in turn has been influenced by Frisch's work on decision models. Briefly my view can be expressed as follows: in his discussion of the relative merits of alternative policy instruments, Lundberg largely limits himself to a qualitative presentation. He gives the reader a fairly exhaustive catalogue of advantages and disadvantages of each instrument, but since none of this is in quantitative terms it is difficult, if not impossible, to make a choice on this basis. Or rather, each person will make a choice according to his own ideological predilections. Since so much of the debate, in Sweden at least, refers to means rather than ends it follows that progress on the econometric front would reduce the area of disagreement to real conflicts of interest. While Lundberg might agree with this, he would probably add that in our present state of insufficient knowledge he has followed the only practical approach by limiting himself to historical case studies rather than embarking upon the difficult road of econometric research.

But the ambiguity does not end here. Even if we admit our ignorance about numerical relationships we must still, for the sake of clarity, specify the number of goals before anything sensible can be said about the number and types of means which must be applied in order to reach the targets. While Lundberg, as mentioned, presents a classification of targets according to their degree of ambition, his discussion of means is not sufficiently related to this classification. Thus it is not possible to appraise, say, the relative merits of monetary and fiscal policies without relating them to some specific targets.

To pursue this subject a little further, we can with Tinbergen formulate the problem of means and ends in a decision model. Four groups of variables enter into the model: (1) *target variables*; e.g., maintenance of full employment, elimination of a balance-of-payments deficit, etc.; (2) *instrument variables*; e.g., certain tax rates, foreign exchange rates, or banking policies; (3) *data*; i.e., parameters given from the "outside," e.g., world market prices; and (4) *irrelevant*

---

4. See J. Tinbergen, *On the Theory of Economic Policy* (Amsterdam: North-Holland Publishing Company, 1952).

variables; i.e., such variables which must be included in a realistic model but which are without political interest either as ends or as means. For the rest, the model resembles the traditional ones: it has definitional equations, technological relationships and behavior equations. In general, the problem then becomes one of examining the conditions under which a given set of target variables have a solution. In a well-defined model where the number of target variables is equal to the number of instrument variables a unique solution may be found. If the number of available instruments exceeds the number of targets, an infinite number of solutions exists. In this case the excess number of degrees of freedom may be eliminated either by adding to the number of targets or by dropping superfluous instruments. On the other hand, if the number of targets exceeds the number of instruments, this implies that some targets must be eliminated or else that some new instruments must be added.

The main purpose of models of this type is, of course, to serve as a basis for econometric work. As one feature of this construct we can also derive a precise definition of the efficacy of an instrument variable: it can be expressed as a partial derivative relating the change in an instrument variable to the resulting change in some particular target variable. It must be added, however, that in the realm of practical policies certain boundary conditions often have to be imposed on the variations of certain variables. For example, a fall in money wages can only take place within narrow limits. Formally, this is done by adding the boundary conditions as a set of inequalities.

Although none of this formal analysis is to be found in the book, at least not explicitly, it seems to provide a convenient analytic framework for policy discussions. In relation to this framework Lundberg's general view can be expressed as follows: many instruments, but in particular specific controls, are burdened with serious imperfections which limit their efficacy. And further, in those cases where the success of a policy depends upon correct forecasting we should go easy because of the poor record of the forecasters. For both of these reasons he therefore advocates limited targets, i.e., objectives which are realizable with relatively simple control measures. I shall come back later to the positive content of his program, but first it is necessary to look closer at his appraisal of some key instruments.

IV

Fiscal policy should, according to Lundberg, be assigned a fairly modest role in cycle policy. In particular he argues that some of its supposed advantages over monetary policy are imaginary rather that

real. For one thing the apparent simplicity of budgetary policy as a control of cyclical fluctuations is due to an oversimplified aggregative analysis. So much depends upon specific forms of taxes and expenditures that the use of a few total categories or propensities may be very misleading. While this may all be true he apparently forgets his own advice on one occasion; in dealing with the case of a balanced increase of taxes and public expenditures he makes the unqualified statement that such a policy is expansionary. The answer is, of course, that the balanced-budget theorem is only correct under very special conditions concerning primary and secondary effects.

Lundberg argues further that fiscal policy is not as precise a tool as is often believed. On the tax side it is not possible to legislate receipts directly, but only the rate structure. And both for taxes and expenditures it is true that indirect as well as more direct effects must be taken into consideration. Consequently it is not obvious that greater precision is achieved than with monetary tools.

But Lundberg has still another weapon in his arsenal. A major point is his emphasis on our imperfect forecasting techniques. Unfortunately limitations of space do not permit me to summarize his interesting technical analysis of the concepts of inflationary gaps and national budgets against the background of postwar Swedish experience. He is undoubtedly right in pointing out that the record of the national budget technique has not been too impressive. This view is confirmed by a recent study undertaken by the Economic Commission for Europe which deals with the experience of the Scandinavian countries, England and the Netherlands.[5] It must be said, however, that one reason the results have been unsatisfactory in Sweden can be found in the fact that not sufficient use has been made of all available information. On the whole it appears that forecasts are based simply on a few summary figures for past rates of growth. In Denmark and Norway, on the other hand, an attempt has been made to estimate the contributions of individual sectors of the economy on the basis of such factors as available raw material and labor supplies, rationing, long term contracts, etc. As a result these forecasts have been better than their Swedish counterparts. In comparing forecasts with actual results an interesting observation can, by the way, be made: on the whole the national budget forecasts have tended to fall below the actual performance of the economy during the period of inflation and growth. What is not clear is whether this is due to political considerations or simply to the cautiousness of the forecasters.

5. *Economic Bulletin for Europe*, first quarter, 1953.

I have mentioned that Lundberg rejects welfare constructions in dealing with policy questions. Nonetheless, when arguing against too much reliance on fiscal policy he does introduce such considerations. His reasoning is, briefly, that in an economy with direct controls and repressed inflation, prices are no longer true indicators of relative values, so that there are difficulties in the way of taking social profitability into account. For example, how is it possible to appraise a policy which aims at a reduction of consumption in favor of investment if relative prices of these two categories are out of line with true demand and cost conditions? For the same reason he also concludes that it is difficult to interpret changes in national income under conditions of disequilibrium.

While much of what Lundberg has to say about the limitations of fiscal policy is an expression of sound scepticism, he does seem to overstate his case. In the first place fiscal policy can be a powerful supplement to monetary controls. Especially in situations of major imbalance an exclusive reliance on monetary means seems less preferable than a flexible use of both instruments. Or take a situation where, under full employment, public expenditures for some reason must be expanded. In order to prevent the secondary effects of such an expansion it seems that fiscal rather than monetary measures must be taken. Secondly, while it is true that our forecasting methods are imperfect, this particular point does not affect, or course, the importance of built-in-stabilizers. Thirdly, on the use of prices as guides for policy decisions, Lundberg is perfectly correct in emphasizing their limited value in a situation of suppressed inflation. However, in recent years the relaxation of controls has been widespread, both in Sweden and elsewhere, and prices are now much more reliable indicators of social values. Still existing deviations are primarily due to imperfections in private markets. While this situation facilitates the formulation of fiscal policy ends, it must at the same time be admitted that in a system of free markets national budget forecasts become more difficult to undertake. It is probably in part for this reason that, with the cessation of postwar inflationary pressures, the national budget technique has lost some of its previous popularity. Finally, and in line with a previous argument, it must be stressed that the role to be assigned to fiscal policy cannot be decided *in abstracto*. That role depends very much on the economic targets; or, put another way, a predominant reliance on monetary policy implies a reduction in the number of possible goals.

V

The following few remarks are intended to reproduce the gist of Lundberg's theoretical argument about monetary policy; unfortunately they cannot touch upon his interesting survey of Swedish policy and thinking before and after the last war. To begin with the conclusion, he calls for a policy which as its theoretical underpinning has a synthesis of a modernized quantity theory, the Keynesian liquidity preference theory, and the Wicksellian analysis of the relationship between the rate of interest and the volume of investment. As far as the quantity theory is concerned, he agrees, of course, with the modern criticism which has centered on the instability of the velocity of circulation. On the other hand, while admitting the validity of this criticism, he maintains that there has been a tendency until recently to neglect the direct importance of total liquidity and its distribution within the system. In this respect he feels that the Keynes of *A Treatise on Money* has more to offer than the Keynes of the *General Theory*. In particular he finds value in the concepts of "availability of credit" and "the fringe of unsatisfied borrowers." Lundberg does not seem to reject the liquidity preference theory, but he is of the opinion that this is a very special way of introducing money and that it excludes any direct impact which the liquidity position may have on consumption and investment expenditures. Finally he feels, as a good Swede, that the Wicksellian analysis of cumulative processes should be a cornerstone in monetary theory.

A consequence of this approach is that it becomes impossible to make any sharp distinction between the quantitative and interest rate effects of monetary policy. Take a situation when the central bank tightens the credit market through a rise in interest rates. In this case it is not only a question of potential borrowers taking less credit; it is also a situation where less credit is likely to be given, that is, the fringe of unsatisfied borrowers will increase unless the more restrictive policy leads to a postponement of new investment projects. A harder credit climate will thus lead to a reappraisal of existing plans and to marginal changes in the propensity to spend.

A precise determination of how, and by how much, a change in monetary policy will affect the economy is not possible. Although this is a weakness of this instrument it can also, according to Lundberg, be a strength. Under favorable conditions relatively small measures may suffice if the monetary authorities can succeed in influencing expectation about the future. This possibility is in part based upon the uncertainty which, say, an increased interest rate will bring about. In the first place such a move may, because of fall-

ing bond prices, temporarily reduce the importance of bonds as secondary liquidity reserves. It will only do so, however, if the market expectation is that present interest rates are abnormally high and that a downward revision eventually will take place. Secondly such a policy may raise doubts in the business community about future possibilities of raising funds. A modest increase in rates, which does not in itself reduce present profit margins a great deal, may nonetheless act as a deterrent on new investment loans. Thirdly, the tightening of credit may be interpreted by business as a warning that the boom is coming to an end. In such a case the monetary authorities are operating on expectations about future profitability.

It should be added that in dealing with monetary policy Lundberg is very much aware of, and emphasizes, the imperfections of the markets for money and capital. (It is, of course, implicit in the concept of a fringe of unsatisfied borrowers). His own view is that in spite of all imperfections there is enough competition among lenders and borrowers to make the arrangement workable. Over-all control of credit is important but it should be based upon methods which are as general as possible. That is to say, the policy should be guided by such broad goals as the levels of prices and employment, or the balance-of-payments position, and it is as a rule undesirable to use the credit instrument to achieve detailed allocation of resources. In fact, he argues that a rationing of credit according to uses partakes of many of the disadvantages of direct physical controls.

VI

In one of the final chapters in the book Lundberg discusses the problem of wages under full employment against the background of postwar Swedish experience. The importance of the question is seen from even the briefest examination of the facts: hourly wages in industry have more than doubled since 1945, a rate of increase which has greatly exceeded the growth in productivity. The general features of the explanation of the sequence are well known. The crucial link is the high level of profits associated with full employment. Such a situation will, on the one hand, induce trade unions to demand higher wages. At the same time the resistance of employers to these demands may not be too great. Profits are such as to permit a higher wage bill, and in monopolistic industries producing for the home market it is usually possible to shift the additional wage cost to the consumer. Employers may also be afraid of losing labor to other firms or industries which are prepared to meet the demands of organized labor, and finally the incentive to resist may be weakened by

high marginal tax rates. In a country like Sweden the process is not limited to wages. Other important groups, such as farmers, civil servants, and social security beneficiaries, have their prices or incomes determined by the cost-of-living index, i.e., indirectly by the level of wages.

For limited periods trade unions and political leaders may enlist the co-operation of labor in a stabilizing policy, but the situation is essentially an unstable one in view of the high level of profits and the strengthened tactical position of the trade unions. Appeals to "reason" and "responsibility" are thus unlikely to work for very long. It is possible, however, that unions after repeated efforts to improve the standard of living of their members find that the beneficial effects of higher wages are nullified by higher prices, but the strength of this factor is weakened by the fact that it will always be in the interest of individual unions to achieve higher wages. Thus the argument presupposes a high degree of solidarity within the trade union movement. There is also the further danger that a degree of sophistication is reached where the unions, when they realize that higher wages will lead to higher prices, increase their demands in order to achieve a given increment to real wages.[6]

Lundberg's proposed solution, which he puts forward in a polemic with G. Rehn of the Confederation of Swedish Trade Unions, is to restrict effective demand through monetary measures until profits are reduced to a point where there is no longer any room for wage increases. While probably nobody would argue against the elimination of excess monetary demand, the crucial question is whether this will suffice to bring about wage and price stability. It is still possible, as Lundberg realizes, that wages may continue to rise even if overall demand has been brought into balance with supply if profits are high enough to allow some room for wage increases. But that means, and he makes this quite explicit, that his proposal will imply some unemployment. Perhaps the greatest weakness of his case is that he does not try to estimate the cost of wage and price stability in terms of unemployment. He does emphasize that our empirical knowledge is limited in this field and that there does not seem to exist any simple mechanical relationships between wages on the one side and profits and unemployment on the other. But in pointing to these uncertainties he does, it seems, reduce the appeal of his program. It

6. In an interesting piece of analysis Lundberg deals with this problem. He defines a wage multiplier which indicates the wage increases required to offset a given rise of prices. This analysis is summarized fully in a review of the book by Hans Brems (*Journal of Political Economy*, June 1954).

may be asked, and with some justification, whether the cost of wage stability will not be too high, and whether some other approach may not be better.

One such proposal has been put forward by the Swedish trade union movement. The central idea is here again to reduce profits in order to prevent the profits-wage spiral. However, according to this program the narrowing of profit margins should be achieved through indirect taxes. Apparently Rehn is prepared to go farther than Lundberg on the restrictive side, but there is this important difference that the trade union proposal envisages rapid use of government resources whenever the restrictive tax policy threatens to create unemployment. In other words, whenever islands of unemployment occur, the plan calls for a differentiated fiscal policy on the expenditure side. Conversely, when inflationary pressures reappear, the Treasury should sterilize tax receipts.

Lundberg expresses various doubts about this program. Thus he questions the assumption that wage demands will be moderated when profits are transferred to the Treasury via indirect taxes, especially since the plan implies a guarantee of full employment in all sectors of the economy. And further, and this is in line with his general view on fiscal policy, he doubts that the government is capable of using this instrument with any great precision. Finally he fears that the siphoning off of profits to the public sector may lead to a misallocation of resources.

While this whole discussion is rather illuminating, and does not seem to have as detailed a counterpart in the Anglo-Saxon literature, it can safely be said that this challenging problem still remains largely unsolved.

VII

Reduced to its essentials Lundberg's program can be summarized as follows. In the first place he wants to establish over-all balance in the economy through monetary policy supplemented by fiscal measures. He wants to do so in order to create conditions favorable to the working of the price mechanism. In the second place he wants to take certain positive steps designed to strengthen the market mechanism. They are measures which (i) increase the mobility of labor and other factors of production (retraining, housing reserve, contribution to transfer costs, etc.); (ii) guarantee a sufficient supply of risk capital; and (iii) increase consumer guidance.

His faith in a program along these lines is in the final analysis an expression of a *Weltanschauung*. This is so not only because differ-

ent people have different values, but also because our knowledge of the working of various policy instruments is still incomplete. One of Lundberg's greatest contributions in the present volume is to explore the relative advantages and disadvantages of alternative instruments, but he would probably be the first one to admit that the evidence he presents is not conclusive. He has built up a case against direct controls in particular, but it is important to keep in mind that in doing so he is limiting himself to situations where only partial and *ad hoc* controls are used. What he is criticizing is a system limited to half-measures in the field of controls, and not a more comprehensive arrangement where the various controls are integrated and co-ordinated.

But to return to his own program it can be argued that his analysis is incomplete. For one thing he does not deal adequately with the social cost of his positive proposals. It may be that the advantages are on the side of, say, a housing reserve rather than government control of the location of new firms or the regional pattern of public expenditures according to labor availability, but the question should at least be explored. Similarly, a sufficient supply of risk capital may call for a degree of income inequality which is in conflict with other social considerations. Moreover, and this is perhaps a more serious objection, while he does protect himself by listing various imperfections of the price system, he evidently does not attach too much importance to them when it comes to the development of his own program, which after all relies heavily on the efficient working of the market mechanism. Are monopoly elements fairly unimportant in the Swedish economy? Are discrepancies between social and marginal productivities moderate? Is the rate of investment the "optimal" one? These and a multitude of related questions have received only a partial treatment. However, implicit in Lundberg's general view on economic policy is the position that market imperfections are tolerable and do not justify a comprehensive system of controls.

In spite of the various critical remarks with which these notes have been intermingled, it must be said, in conclusion, that the author has produced an immensely readable and interesting study in public policy. It is to be hoped that a translation into English will make it accessible to a wider audience.

SVEND LAURSEN.

BRANDEIS UNIVERSITY,
WALTHAM, MASSACHUSETTS

THE PROBLEM OF "UNDERDEVELOPMENT" IN THE ENGLISH CLASSICAL SCHOOL*

By Erskine McKinley

I. INTRODUCTION

The English classical school[1] wrote at a time when agricultural productivity in England threatened to serve as a severe constraint upon the industrialization of that country. They cast the problem of economic "development" in terms of the displacement of an economy from stationary equilibrium, historical or hypothetical, and its movement towards such an equilibrium.

A stationary economy is characterized by a zero rate of population increase and a zero rate of net investment. Given techniques, the return to a unit of agricultural land at the margin produces just that surplus over labor costs which is regarded by the capitalist-farmer as adequate to keep that unit of land in cultivation but as inadequate to induce further intensive or extensive cultivation. The surplus, or profit, may be large in real terms yet still be regarded in some sense subjectively as "minimum"; or it may actually be very small in real terms. Only in the former case may the economy be regarded as underdeveloped; in the latter, it is mature. It must be presumed that if improved techniques are known but not introduced, they have not served to elevate the schedule of the marginal efficiency of capital enough to induce new investment. The difference between the subjective and objective nature of the rate of profits accounts for the fact that the "observed" minimum rate of profits may vary from economy to economy.[2] China, for example, was cited as a stationary

1. This school is defined to include Adam Smith, Ricardo, Malthus, McCulloch, Senior, and the two Mills.

2. Cf. Lionel Robbins, "On a Certain Ambiguity in the Conception of Stationary Equilibrium," *Economic Journal*, June 1930, esp. pp. 198–200.

economy with a high rate of profits. Adam Smith,[3] McCulloch,[4] and J. S. Mill,[5] all referred to it. But although they agreed to its stationariness, Smith attributed the absence of net investment to monopoly and government interference; McCulloch, to high risk; and Mill, to the overvaluing of present as against future pleasures. Mill also imputed underdevelopment, if not actual stationariness, to Asia, Russia, Hungary, Spain, and Ireland.[6] Ireland was taken as an example of stationary underdevelopment also by Malthus and Ricardo. They referred as well to New Spain.[7]

Since net investment in the classical school meant an advance of wage goods to the labor force which served to raise the rate of real wages above the level of subsistence, a rate of population increase of zero implies that the laboring population is at subsistence whether it be held there physiologically or psychologically. Consequently, the crucial variables for lifting an economy out of underdeveloped stationariness are the level of net investment and the rate of population increase.

II. THE PROBLEM:
CAPITAL ACCUMULATION AND POPULATION GROWTH

Population will be mentioned only briefly. Even contemporary demographers disagree in their diagnoses and recommendations for underdeveloped areas. With regard to capital, the classical attitude towards (1) its source or supply, (2) the effective desire of accumulation, or the demand for capital, (3) the logical directions of investment for development, or, as one might say today, investment priorities, (4) the rate of investment that is optimal, requires discussion. Only at this point will it be necessary to enter briefly into the classical position on the possibility of factor substitution.

Population: the shift in the classical school.

Malthus' first *Essay* on population posited a principle that is completely physiological, i.e., a vast and constant stream of births feeds a population which must be "tailored" to its physical environment by a death rate which is governed basically by the scarcity of

3. *Wealth of Nations*, Modern Library Edition, p. 72.
4. *Principles of Political Economy* (Edinburgh, 1825), p. 105.
5. *Principles of Political Economy* (London, 1848), I, 128.
6. *Ibid.*, p. 226.
7. Malthus, *Principles of Political Economy* (New York: Augustus Kelley, 1951), pp. 336–41, 345–46 *et passim; The Works and Correspondence of David Ricardo*, ed. Piero Sraffa (Cambridge University Press, 1951–52), VII, 184 and *passim.*

food. Ricardo adopted this position, weakening it somewhat by admitting the possibility, though by no means the inevitability, of a secular upward movement of the level of subsistence. Malthus never, I think, really forsook his early position except to the extent of admitting the possibility, though here again never granting its inevitability, of there arising preventive checks in the course of economic development. One finds in the *Essay* not much hope that such checks will be automatically induced. It is true, however, that in the *Principles* he appears much more sanguine.[8]

The classical school couched the question they put themselves in terms of whether there is a "natural" tendency for capital accumulation to outstrip population increase. It is curious that James Mill and J. R. McCulloch reached totally different conclusions from their observation of the same phenomenon. Mill, impressed by the extent of poverty in the world, concluded that there was no such "natural" tendency. McCulloch, impressed by the world's accumulated riches, concluded that there was.[9] The real crux of the matter, we would say today, is not so much the "naturalness" of any tendency but rather the automaticity in the course of economic development of there evolving "a cultural pattern suited eventually to bring population growth to a stand."[1] While not one member of the classical school was completely consistent in this regard, the consensus, after the death of Ricardo, seems to have been that such a cultural pattern would be induced.[2] They would seem, at least at this juncture, to

8. J. J. Spengler argues very convincingly that the *Essay* reveals the causes of poverty; the *Principles*, the causes of its mitigation. See his "Malthus's Total Population Theory," *Canadian Journal of Economics and Political Science*, XI (Feb. and May, 1945), 83–110, 234–64. Senior apparently felt that Malthus, if not abandoning his early position, abandoned at least its more dire results. On balance, I suspect he was mistaken. But cf. James Bonar, *Malthus and His Work*, who more agrees than disagrees with Senior; Marian Bowley, *Nassau Senior and Classical Economics* (London, 1937), pp. 118 ff.; and Kenneth Smith, *The Malthusian Controversy* (London, 1951), pp. 184 ff., both of whom more disagree than agree with Senior.

9. James Mill, *Elements of Political Economy* (3d ed., 1826), pp. 44–45, and McCulloch, *op. cit.*, p. 18.

1. Spengler, *op. cit.*, p. 100.

2. "Far from population having a tendency to increase faster than subsistence, Senior and Tooke held that there were facts enough to prove the reverse was the case; and as to McCulloch, who is always bitter against Malthus, the workings of an envious and mean disposition, he held that there was in human nature a principle of improvement and exertion that was at all times sufficient to counteract and overcome the principle of population, and therefore that Malthus' theory was altogether erroneous, or as he expressed it in his own happy language, false. Torrens likewise attacked the principle. . . ," "From J. L.

have been right. However, there is concern expressed in some quarters today lest any attempt to foster industrialization be overborne by a population explosion, and some doubt as to the transferability of the widespread habit of birth prevention. In any event, the classical school were aware that capital could increase with rapidity but with no repercussions on welfare if that capital increase were matched by an increase in population. And some of them, even granting automaticity of the change in cultural patterns leading to a reduction in the birth-rate, would certainly have approved of positive measures lest the *lag* in cultural adjustments cause any attempt at industrialization to abort.

Capital: its source or supply.

It is possible to distinguish four means of obtaining capital in the classical model. Two of these, the appropriation of rents and inflation, involve the rechanneling or redistribution of existing incomes. Two of them, land reform to eliminate "disguised unemployment," and the evoking of an increased intensity of effort, involve the creation of new capital within existing techniques.

It might be inferred from the Ricardian model that in a stationary state there must be an extreme attenuation of the distribution pyramid with a comparatively small number of landowners receiving a large proportion of the national product. This may indeed be true but it should be pointed out that there is no necessity for such attenuation. For, as James Mill indicated,[3] if the progress towards stationariness has been accompanied by active sales or other legal division of land, there may well have arrived a situation in which there are many landowners each with so small a plot that they are no better off as regards per capita income than other income groups. Further, as Senior made explicit what Ricardo strongly implied, not only the landowners but most individuals in an economy receive incomes in excess of the absolute minimum necessary supply price. These incomes, even when received by the wage-earner, Ricardo

Mallet's Diaries," Political Economy Club, *Centenary Volume* VI (London, 1921), 265–66. See also the statement of a further consensus that "One of the errors of Ricardo seems to have been to have followed up Malthus' principles of population to unwarrantable conclusions," p. 225. Mallet's urbanity, perceptiveness, and occasional waspishness, make the reading of these extracts, ever too brief, a rare delight.

3. *Op. cit.*, pp. 52–53.

called "profits."[4] Senior called them "rents" or "profits." In the event that techniques whereby output may be increased are known and not employed such rents are the logical internal source of the capital necessary to introduce the innovation. The great difficulty is to isolate the element of "pure rent" for if more than this is infringed, supply of the factor will be curtailed with a consequent diminution of output. And even if the rents can be identified, there remains the further difficulty of aggregating them. This may simply be a matter of creating the appropriate credit institutions such as savings banks. Provision of such institutions may, indeed, lead to the automatic inducement into productive channels of otherwise hoarded or squandered funds.[5] Where such hoards and squandered incomes bulk fairly large individually there remains still the problem of why advantage has not been taken of the innovation earlier, in other words why is the entrepreneurial spirit lacking in the economy. This question will arise again.

Accustomed as we are to the cliché that the classical economists thought only in real terms, ignoring money, it may come as something of a surprise to find that they were quite familiar with inflation as a means of capital accumulation. This they saw to operate through favoring the entrepreneurial class to the hardship of the wage-earner and annuitant (or, sometimes, as favoring the first two classes to the detriment of the third). Thus Ricardo, feeling that the entrepreneur alone would benefit by the "depreciation of money" for the purpose of accelerating capital accumulation, rejected the method as augmenting riches only by "diminishing happiness."[6] Malthus, too, explicitly rejected inflation as a method but for a somewhat different reason. Probably, to be sure, he did disapprove of the inequities which inflation works in its course, but his chief grounds was that he feared, in its being carried to excess, the danger of a cyclical downturn of such severity that "landlords, capitalists, and labourers . . . may

---

4. Sraffa, I, 348 and 348, n. Also, " . . . the wages themselves may be considered as part of the profits of stock, — and are frequently the foundation of new capital." *Ibid.*, VI, 147.

5. Thus, recently we read that in Turkey "the peasant is taking his money out of the mattress and investing it in something other than gold trinkets for his wife. . . . Although the Turkish domestic markets for equities are still in their infancy . . . their growth . . . is . . . one of the most significant developments since Turkey abandoned the policy of state-operated industry in 1950." *New York Times*, Jan. 24, 1954, p. 11.

6. Sraffa, VI, 233.

well wish they had not been subjected to the stimulus."[7]   The question came up for discussion at the Political Economy Club.   Mallet reported that during the discussion McCulloch, who had none of Ricardo's or Malthus's fears,

in his sarcastic and cynical manner derided Mr. Tooke's concern for old gentlemen and ladies, dowagers, spinsters and landholders.   He cared not what became of them, and whether they were driven from the parlour to the garret, provided the producers — the productive and industrious classes — were benefited, which he had no doubt they were by a gradual depreciation in the value of money.[8]

J. S. Mill did not approve of long-continued forced saving but he did point out the need, implied in all the classical writings, of credit for rendering fixed wealth mobile.[9]

Land reform in underdeveloped areas was a universal prescription of the classical school.   That was necessary only where custom had resisted the division of land into the most efficient production unit.   Thus, Malthus recommended the breaking up of large landed estates.[1]   He cautioned, however, against too great an atomization of the farm unit, which he felt would militate against the great advan-

7. *Op. cit.*, p. 168.
8. *Op. cit.*, p. 219.
9. *Op. cit.*, I, 427 and II, 38.   Mill did not go so far as Senior, according to Mallet who reports that " . . . both Mr. Palmer and Senior and the Rev. Sidney Smith, who was our visitor, held that the prosperity of the country (if it be prosperous) was owing to paper money and the facilities of credit afforded by banking.   They even went so far as to describe issues of paper money as a *creation of capital*, and Sidney Smith talked of the moral influence country banks had: stimulating industry, encouraging exertions and good character, and so forth; of which anything more incorrect cannot well be conceived; for I take it that the facilities of credit given by country Banks have made ten speculators for one industrious man, and have given a feverish and artificial stimulus to business and agriculture, of which the effects have been anything but salutary. . . .   All that a Bank can do — I mean all the good of which it can be the instrument to the community — consists in collecting in the way of deposits accumulations of capital obtained at a low rate of interest, and which would not otherwise be beneficially employed, but might be hoarded or wasted in unprofitable speculations, and distributing these . . . to those who are in want of capital — to the industrious and skillful."   *Op. cit.*, p. 240.   I think the novelty of this belief of Palmer, Senior, and Smith, that credit facilities can lead to the "creation" of real capital in any sense other than by forced saving can be overemphasized.   After all, they were saying that the actual *source* of new capital was "stimulated" industry and "encouraged" exertions.   Mallet, on the other hand, underemphasized the *positive* effects — as distinguished from the neutral — on capital formation that may result from the distribution of otherwise squandered or hoarded funds to the "industrious and skillful."   Cf. Hayek's "A Note on the Development of the Doctrine of 'Forced Saving', " in his *Profits, Interest and Investment.*
1. *Op. cit.*, p. 373.

tages to be reaped from the division of labor.[2] On this basis, Ricardo, writing to Trower about the condition of Ireland, mentioned in passing that "small farms would be laid into large,"[3] this in the early stage of beginning development. "Disguised unemployment" in the modern sense is perhaps too much to infer from so casual a remark but Arthur Young observed on the continent, and his observation was quoted by J. S. Mill with approval, just such a condition. Thus Mill said that the division of land might be carried so far "as not fully to occupy the time and attention of the family."[4] This quite clearly implies to me a situation such that hands might be transferred from agriculture into industry and, though total food production might not be actually increased it would, at the least, remain undiminished.[5] But here again we have no prescription for how the reform is to be accomplished, much less how the displaced laborer is to be got into industry and the food he formerly consumed channeled back into his possession. McCulloch, somewhat inconsistently with his position on inflation cited above, further cautioned that impoverishment of the landlord in the process of land reform (or suddenly indulged habits of parsimony) might "destroy one of the great incentives to industry and powerful intellectual exertions . . . [since] . . . many a fortune originated in the desire of living like the great of the land."[6]

This leads into a further consideration of the classical school: that there exists a considerable pool of potential capital which can be tapped by simple increased exertion or greater intensity of effort. Just so, since the introduction of cost-reducing innovations in the production of the basic staples may encounter an inelastic demand for the increased output (if population growth does not absorb it) there

2. *Ibid.*, p. 376.
3. Sraffa, IX, 153.
4. *Op. cit.*, I, 326.
5. It might, I think, be said that the unanimous opposition of the school to the Poor Laws was based on a kind of underemployment conception. They felt that the Poor Laws by supplementing submarginal wages to or above the margin, particularly as the rates were based on family size, encouraged the growth of a population beyond its economic justification. This was the sense in which Ricardo and James Mill meant that capital accumulation if continued beyond the economically correct point would result in impoverishment of all classes. Thus Mallet said that " . . . we see that in Ireland, where rent is absorbing everything, in consequence of the immense competition for land, a system of Poor Laws is likely soon to equalize the division." *Op. cit.*, p. 225.
6. Reported by Mallet, *op. cit.*, p. 264. There are echoes here of Malthus' position in deploring actions which might curtail demand and lead to a premature stagnation of the economy.

must be demand-increasing, as well as cost-reducing, innovations.[7]
Malthus especially was dubious of the development-promoting effi-
cacy of cost-reducing innovations since he felt that the worker might
well spend his increased income on "indolence."[8] And it is quite true
that although the wealthy, through the operation of the demonstra-
tion of new goods, might be expected to switch from the demand for
menial services to that for material luxuries or to have goods pro-
duced to exchange for exotic imports, the working classes, having no
surplus income to shift, *must* be led to increased effort. Mill even
expressed the germ of the Toynbee thesis that obstacles to desire
must exist in just the proper degree of stubbornness to evoke such
increased effort.[9] We might even go so far as to read into McCulloch
the idea that taxes may be beneficial in providing some such necessary
stimulus. Thus, Mallet reported of an evening at his Political
Economy Club:

Some things struck me this evening: 1st. The decided opinion which appeared to
prevail that taxation and even heavy taxation, under such circumstances as those
of the French War, very much tend to stimulate industry, and had in fact led to
a much greater amount of savings and accumulation, than would otherwise have
been the case. . . .[1]

7. Viewing the matter in retrospect, Warren C. Scoville says: " . . . one
must infer that Mr. Brozen thinks purely cost-reducing innovations and imita-
tions are quantitatively more important than hybrid or purely demand-increasing
innovations. This may well be the case although some economists may be prone
to question the conclusion on empirical grounds." *American Economic Review*,
Proceedings, XLI (May 1951), 277.

8. Ricardo agreed in his correspondence however little he may have stressed
this in the *Principles*. Thus, very cheap food, in Ireland the potato, in New
Spain manioc and the banana, available with little effort, account for the under-
development there. He and Malthus seem to have agreed that the cause for the
underdevelopment lay in what we would call today a sharply backward-bending
supply curve of labor, or in a highly inelastic schedule of the demand for income
in terms of effort. See Malthus, *op. cit.*, pp. 336–41 and *passim*. Also, Sraffa, VII,
184, and cf. McCulloch, *op. cit.*, pp. 396–98.

9. *Op. cit.*, I, 122–23.

1. *Op. cit.*, p. 226. Mallet went on to add, and it does not bear out his
opinion that taxation leads to a *greater* amount of savings: "McCulloch said in so
many words, *that not a farthing more would have been accumulated; that there would
not be a farthing more capital in the Country if the National Debt had not been incurred;*
which seems to be generally assented to." However, later, p. 271, he says: "But
to return to McCulloch — he held that the pressure of taxation was the greatest
of all incentives to industry, and during a war, when taxes come pelting in, it
stimulates the industrious man to greater and greater exertions; so that 20 years
of war were more likely, according to his view, to benefit the country and add to
its wealth, than 20 years of peace."

The school saw clearly enough that there must always exist a lag between demand and its gratification and that if an increment to income achieved by increased effort is consumed, no saving can take place. Mill, however, seemed to have believed that the act of saving for the sake of gratifying some consumption desire would lead to a *habit* of saving which would continue even after the objective was achieved.[2] The assumption has been, of course, either that economically warranted resources are available under given techniques or that innovations for increasing output are available. If neither is true, reallocation of income into the hands of the labor force, if it induces further population increase, will simply result in the impoverishment of the expropriated class with no balancing amelioration in the condition of any other class. Thus, for the growth of capital James Mill very boldly suggested an income tax and channeling of the proceeds into entrepreneurial hands. He rejected his own suggestion because, Malthusian that he was, he felt that the effects would be only leveling. Exactly what so-convinced a utilitarian might have thought of this scheme if the existence of innovations at hand were sufficient to produce a virtual explosion of capital formation can only be conjectured.[3]

Capital: the demand side — the effective desire of accumulation.

B. S. Keirstead makes the rather surprising statement that "the orderly and automatic adjustments of Ricardo's market leave no function for the entrepreneur to perform. He has no problem of anticipation, no real risk to bear, no serious charge is put upon his judgment."[4] While it is true that in the *Principles* Ricardo did not discourse at any length upon the duties and functions of the entrepreneur, a not inconsiderable role in the system is assigned to him implicitly. In the first place, he is expected to innovate and the entrepreneur who does not himself innovate is expected to imitate the innovator with remarkable speed. It has been said that the "economic man" does not move of his own volition but gravitates. Ricardo, however, indicated that frictional disorder can be painfully prolonged by an understandable inertia on the part of the entre-

2. *Op. cit.*, I, p. 119.
3. *Op. cit.*, pp. 60–61. In fairness to Mill I must add, lest it be objected that there is nothing intrinsically evil about "leveling," that his fear was that there could be no cushioning of natural disaster should it fall upon a population whose every member was at the level of poverty and barest subsistence.
4. *The Theory of Economic Change*, p. 81.

preneur. The discomforts of the adjustment to structural change are thus prolonged by the "opposition which is naturally given to such change," and the speedy reaction to changes in market indicators depends on "the degree of information, or the strength of the prejudice of those who offer it . . . ." Perhaps for yet unborn econometricians he added: "it cannot be the subject of any thing like accurate calculation."[5] And in the *Principles* he said that the reallocation of capital occasions distress which "will be longer or shorter according to the strength of that disinclination which most men feel to abandon that employment of their capital to which they have been long accustomed."[6] I agree that the function of the entrepreneur may be a gravitational one — the difficulty is that he must know when, how fast, and in what direction to gravitate. It is *difficult* to be an economic man and this has implications for the underdeveloped areas.[7]

Ricardo, Malthus, and the rest of the school saw the entrepreneur as uniquely the economic actor with both the desire and the means to accumulate. And just as J. S. Mill spoke of the "habit" of saving on the part of the wage-earner, so Senior considered entrepreneurial accumulation, too, largely a matter of habituation.[8] The conditions which will induce the growth of this habit are substantially the same as those which will induce increased effort on the part of the labor force. Security in the possession of property is a most important precondition of entrepreneurial activity.[9] McCulloch even listed security as the first of his three conditions for an economy's emergence from barbarism.[1] But this is a precondition and it is

---

5. Sraffa, VII, 66–67.

6. *Ibid.*, I, 265.

7. Cf. P. T. Ellsworth, "Factors in the Economic Development of Ceylon," *American Economic Review*, Proceedings, XLIII (May 1953), 117: "The Ceylonese peasant is notoriously conservative and guided by tradition; he dislikes change and is not easily moved by economic considerations to alter his customary ways, being far from an 'economic man' in the Western sense."

8. *An Outline of the Science of Political Economy* (New York: Augustus Kelley, 1938), p. 192. "Capitals are generally formed from small beginnings by acts of accumulation which become in time habitual."

9. Senior perhaps quite rightly called this a noneconomic factor but did not hesitate to say that "Ireland is physically poor because she is morally and intellectually uneducated. And while she continues uneducated, while the ignorance and violence of her population render persons and property insecure, and prevent the accumulation and prohibit the introduction of capital, legislative measures, intended solely and directly to relieve her poverty, may not indeed be ineffectual, for they may aggravate the disease. . . ." *Ibid.*, p. 135.

1. *Op. cit.*, p. 73. The second is an exchange economy and the third the accumulation of capital. Cf. p. 112.

again McCulloch who I think summarized the attitude of the school when he said:

National prosperity does not depend nearly so much on advantageous situation, salubrity of climate, or fertility of soil, as on the adoption of measures fitted to excite the inventive powers of genius, and to give perseverance and activity to industry.[2]

In the earlier stages, at least before industrialization gets well under way, Smith was quoted by Malthus as pointing out the desirability of a shift in taste from personal services to material luxuries. Ricardo also felt it desirable that released menials should produce "lace, china, furniture," as they came, by the induced change in tastes plus a sense of security, to be demanded.[3] But Malthus warned that "most men place some limits, however variable, to the quantity of conveniences and luxuries they will labour for. . . ."[4] New desires, consequently, must be *continuously* generated as the old are gratified. But there need be no qualitative limit to demanded conveniences and luxuries and there would seem to be no limit at all if the desire is for "distinction."[5] This desire for distinction or for "rising in the world,"[6] can manifest itself only in a society where there is vertical social mobility or where, in Senior's words, the "road to social eminence is most open."[7] And the entrepreneurial ability which both permits and is fed by social betterment need not manifest itself only among the wealthy. It was, so far as I know, J. B. Say who first pointed out that

It is commonly requisite for the adventurer himself to provide the necessary funds. Not that he must be already rich; for he may work upon borrowed capital; but he must at least be solvent, and have the reputation of intelligence, prudence, probity and regularity; and must be able, by the nature of his connections, to procure the loan of capital he may happen himself not to possess. These requisites shut out a great many competitors.[8]

J. S. Mill hoped that education would increase the frequency with which entrepreneurial ability appears. Just as he indicated the neces-

2. *Ibid.*, p. 23.
3. Sraffa, VIII, 102–3.
4. *Op. cit.*, p. 355.
5. Senior, *op. cit.*, p. 12. "The most obvious source of distinction is the possession of *superior* wealth." Italics supplied.
6. McCulloch, *op. cit.*, p. 14.
7. *Op. cit.*, p. 27.
8. *A Treatise on Political Economy*, trans. from the 4th ed. by C. R. Prinsep (Philadelphia, 1857), p. 330.

sity for credit to make capital mobile, so did he look to it to make entrepreneurial ability effective wherever it might appear.[9]   His father, too, had stressed heavily the social desirability of a growing middle class.   Malthus defended primogeniture as it forced the younger sons of the landed aristocracy to enter the middle and enterprising class.   But a certain display was considered necessary to entice the able into wishing to rise into it.   It seems reasonably just to say that the desire for material wealth, in the evoking of increased effort on the part of the labor force and entrepreneurial activity among those capable of it, might be regarded as one of the most crucial requirements for the development of an underdeveloped economy.[1]

The directions for investment: priorities.

It was held generally in the classical school that production in agriculture would show historically diminishing returns.   By this I mean that upward shifts in the schedule of returns caused by innovations would be compensated for or more than compensated for by the downward movement along the schedule caused by the increasing demand for agricultural products.   Production in industry, on the other hand, was thought to be capable of indefinitely increasing returns, either because of economies of scale or because of an indefinite supply of inventions and innovations.   This would imply that the terms of trade between agricultural and manufactured goods are turning more and more against the latter.   For this reason the rate of return in agriculture was considered in some sense as governing the

9. Credit is, thus, "the means by which the industrial talent of the country is turned to most account for the purposes of production."   *Op. cit.*, II, 38.

1. The treatment of a demand for luxuries as an entrepreneurial incentive leads into an obvious dilemma since consumption is in a sense an alternative to saving.   I have the impression, however, that it is not so much the matter of a dilemma as of appropriate lags.   Thus, it is clear enough that the stimulus to the working class to improve their condition was the emulation of the rich.   But just as the level of living of the working class was expected to be rising, so too, surely, was the level of the entrepreneurial class.   The crucial shift in demand in a stationary economy was from services to material luxuries.   Doubtless, for the sake of *growth* stripped of any welfare significance the prime motive is desire of accumulation.   Thus, " . . . the view [was] shared by the early nineteenth-century classicists and Marx, that the almost exclusive source of savings was profits.   The workers had no excess income.   The landowning gentry lacked a bent toward thrift.   But the profit-making commercial classes had both.   This heterogeneity in values and aspirations may also explain the alleged paucity of savings in many backward areas in which the richer classes are supposed to lack a bourgeois appreciation of the satisfaction of accumulation. . . ."   Moses Abramowitz, "Economics of Growth," in *A Survey of Contemporary Economics*,

rate of return in manufacturing, but it was realized that the two rates were mutually determining and this realization is perhaps more germane to our present inquiry.

It is obviously to the advantage of the economy to have capital-equipped labor move into that sector which shows increasing returns — in the classical model the industrial sector. But any unit of labor must be equipped with food, or, as they would have said, with capital. Here we encounter the fixity of the input of labor and variable capital (in the classical sense). This implies first that food and labor beyond the needs of simple reproduction of the economy be available; second, that the surplus food be advanced to the laborer until his output is produced; and, third, that the output of the laborer be considered as a satisfactory exchange for the originally advanced food. Thus it becomes increasingly clear why it is that the annuitant and landlord must have a demand for industrial products. The validity of the assumption of food as a constraint on industrialization will be mentioned at the conclusion of this paper.

Adam Smith, under the influence of the physiocrats, advocated as most advantageous to an economy the investment of labor on the land. Malthus agreed to the extent at least of wishing, for nationalistic and sociological reasons, to see a balance between agriculture and manufacturing.[2] The more radically classical attitude was, however, that of Tooke and McCulloch:

Vol. II, ed. Bernard F. Haley, pp. 150–51. If we leave aside the somewhat doubtful assertion that the workers had no excess income then it is true that the landowning class's lack of thrift must be compensated by a high degree of it in the entrepreneurial class. But this thrift must, in order to manifest itself, have something to work on, primarily the desire for *material* luxuries on the part of the landowning gentry. Despite the strand in classical thought that extolled parsimony, it was recognized that industry must accompany the parsimony, and that industry must have something on which to work. At all events, I get the strong impression that a rising standard of living was the desideratum for all classes. We have as a much less workable alternative the opinions of Adam Smith and James Mill that social pressure might render frugality just as productive of emulation by the poorer classes as expensive display. Thus Smith, *Wealth of Nations*, pp. 96–97, said of Holland that it is there "unfashionable not to be a man of business. Necessity makes it usual for almost every man to be so, and custom every where regulates fashion." And James Mill, *op. cit.*, p. 59, expected the legislature or social pressure "to render frugality fashionable, and expense disgraceful." Even Malthus, the most demand-conscious member of the school said "when the capital of a country is deficient, compared with the demand for its products, a temporary economy of consumption is required, in order to provide the supply of capital which alone can furnish the means of an increased consumption in the future." *Op. cit.*, p. 326.

2. Cf. Bonar, *op. cit.*, pp. 224 ff. for a reasoned analysis of Malthus' position.

Tooke did not think that we should ever import more than 3 millions of quarters annually, which is little more than was imported in 1830. But I do not see the grounds of such an opinion, with Russia and Poland and North America open. Malthus evidently thought it a very hazardous experiment, but McCulloch was ready to turn the whole country into one vast manufacturing district, filled with smoke and steam engines and radical weavers, and to set adrift all the gentlemen and farmers now constituting our agricultural population. Nor was Tooke at all startled by this *beau ideal* of a country supplied from the fertile lands of North America and Russian Poland; only that he did not think it likely to be realized.[3]

Ricardo, with his eyes too on the food-constraint on industrial development, would have said that the first area for development in a stationary economy (with innovations available) must be agriculture but that the surplus obtained could then be used in manufacturing and that a mutually feeding growth can then proceed.[4] John Stuart Mill, no more protectionist than physiocrat, seems to have agreed. He said:

> . . . the best chance of an early development of the productive resources of India consists in the now rapid growth of its exports of agricultural produce. . . . The producers of these articles are consumers of food supplied by their fellow-agriculturalists in India; and the market thus opened for surplus food will, if accompanied by good government, raise up by degrees more extended wants and desires,

3. Mallet, *op. cit.*, pp. 233–34. Mallet's attitude is clear.

4. "It is the accumulation of wealth in Agriculture which first gives the notion and the means of establishing Manufactures. Manufactures which in turn become the cause of new accumulations of capital which tend to produce a fresh demand for labour, an increased population, and a greater consumption of agricultural produce. Thus Agriculture is alternately the cause and effect of manufacturing industry." Sraffa, VIII, 102–3. My belief in Ricardo's assignment of agricultural priority in the developmental process may be challenged. I must adhere to it for two reasons. First, Ricardo's emphasis on the repeal of the Corn Laws would indicate to me that he considered the simplest way of raising profits for increased investment within a given institutional framework would be the lowering of money wages which would be possible in an economy more or less at subsistence only by the cheapening of food. But if this is dismissed as an impracticable means for raising profits, Ricardo seems still to have thought it the readiest fashion for raising the welfare of the subsistence laborers. For he replies to Malthus: "But it may be said that the capitalist's income will not be increased; that the million deducted from the landlord's rent will be paid in additional wages to labourers! Be it so; this will make no difference to the argument: the situation of society will be improved and they will be able to bear the same money burthens with greater facility than before; it will only prove what is still more desirable, that the situation of another class, and by far the most important class in society, is the one which is chiefly benefited by the new distribution." *Ibid.*, 1, 424–25. Cf. G. E. Britnell, "Factors in the Economic Development of Guatemala," *American Economic Review*, Proceedings, XLIII (May 1953), 105: " . . . it seems clear that any material improvement in the standard of living in the near future must come about mainly through increased agricultural production."

directed either towards European commodities, or towards things which will require for their production in India a large manufacturing population.[5]

One direction of investment was given high priority by the school, what we might call social overhead capital, of which two examples, educational facilities and means of transportation and communication might be mentioned. Education, they felt, would serve to slow down the rate of population increase but would, at the same time, lead to increased efficiency. Thus Senior referred to "good education" as "a kind of mental raw material."[6] J. S. Mill, as mentioned before, saw a connection between education and entrepreneurial ability and spoke of "intellectual speculation" as "the most influential part of the productive labour of society. . . ."[7] McCulloch, too, saw education both as the source of continuing invention and as "the means for providing for the permanent improvement of the poor."[8] It may be recalled that Adam Smith considered a group of "philosophers" to be one of the fruits of a division of labor but expressed the fear that the same specialization might lead to a loss of general intelligence on the part of the urbanized workman. McCulloch, quite the opposite of an agrarian fundamentalist, surveyed the evidence and reached the conclusion that there is greater gain in education and intelligence from the easy communication which urbanization affords than loss from the particularizing of tasks.[9] Of transport facilities, J. S. Mill held that "good roads are equivalent to good tools."[1] And similarly Senior saw in good transport facilities the means of counteracting the principle of diminishing returns — a hopeful sign for countries without such facilities and which have settled into stationariness because of the operation of diminishing returns.

In sum, I think we might say, the classical school saw the great desirability and indeed necessity of a growing industrial sector in a

5. *Op. cit.*, I, 145–46.
6. And he defined the term as including "Probity, industry, judgment, elementary knowledge, and the other moral and intellectual habits and acquirements to which we give the general name. . . ." *Op. cit.*, p. 220. It may be mentioned again in this connection that Senior considered all this as capital and the material return to it as "profit."
7. *Op. cit.*, I, 53.
8. *Op. cit.*, pp. 118–19, 359.
9. *Ibid.*, pp. 160–61. He asserted, too, somewhat guardedly, that urbanization produces a healthier population. P. 159.
1. *Op. cit.*, I, 219. Notice the *direction* into which the revenue of the rich might be advantageously diverted. "It was admitted on all sides that the application of the revenues of wealthy men to fixed capital such as roads, canals, railways, piers, mills, was advantageous and desirable. . . ." Mallet, *op. cit.*, p. 264.

developing economy. They realized, however, that the growth of this sector was severely limited by the productivity of agriculture and would have recommended first attention to that area. Thus elevating the schedule of returns to land by additional applications of capital, including, perhaps primarily, social overhead capital, might be expected to yield sufficient surplus of food either to feed fixed capital-producing labor directly or to export in exchange for the import of such fixed capital.

The rate of development: the substitution of factors.

Up to this point capital has been considered as composed solely of wage goods, that is to say, capital as the means of *supporting* a unit of labor. This seems to have been Smith's primary meaning. Parsimony meant to him the decision on the part of the owner of a surplus of food not to consume it himself or to assign it to the consumption of menials but rather to "invest" it by paying it out to a laborer occupied in the production of material wealth.[2] Ricardo, emphasizing the increase in variable capital, at first either treated the fixed capital input coefficient as fixed or ignored it. In the third edition, however, he introduced the idea of the possibility of substitution between fixed capital and labor and proposed that labor can be supplanted by fixed capital with which it is in constant price competition. In the event that the introduction of machinery reduces the gross output of an economy, even though the "social dividend" be increased by the substitution, labor may thus become redundant.[3] Here Ricardo is correct in a static context but McCulloch, who accepted the Ricardian logic while denying its applicability to the real world, was correct if progress is assumed. Indeed, short of population increase which may not be desirable, labor-saving innovation is the obvious source of that "surplus" labor which, if equipped with the means of subsistence, can provide the germ of development. The assumption is, of course, that labor which is supplanted in one line of endeavor will be absorbed in another. Senior, like McCulloch taking progress for granted, adopted the notion of price competition between labor and fixed capital and concluded that their relative prices would (and should) determine the

---

2. Lauderdale was critical, not to say captious, towards this idea and countered that the function of "capital" is not so much to support as to *supplant* labor. Each was, of course, perfectly correct, Smith in a dynamic and Lauderdale in a static context. *Inquiry into the Nature and Causes of Public Wealth* (2d ed., Edinburgh, 1819). For an example of his position see his simile, p. 177 n.

3. Sraffa, I, 386-92.

capital intensity of production.[4] The relevance of this to under-developed economies today is clear enough. The classical school would have argued against the too rapid conversion of wage-capital into fixed capital lest, before the fixed capital began to produce, there be experienced a painful shortage of wage goods. Of course, if there is a significant amount of rural underemployment with a large surplus of food this presents no problem. When, however, the creation or increase of the industrial labor force must involve a reduction in the agricultural labor force with consequent diminution of agricultural output, there must be an agricultural surplus available during the period of construction of the fixed capital. And it must, of course, be made accessible to the industrial labor force.

The classical school would have answered the question as to the rapidity with which an economy can industrialize by saying that it depended on the size of the agricultural surplus and the size of the population available to utilize the surplus in new production. The disadvantage here to a stationary and underdeveloped economy is that it is the absence of one of these factors which renders it stationary. But there is a more hopeful aspect. For the school unanimously felt that a country which is not industrialized has a higher degree of flexibility than one which is.[5] This is true since there is little or no "sunk capital" to become obsolete and less chance of technological displacement attendant upon the introduction of new and more efficient techniques of production. This is, I suppose, to underline the obvious: shifts in demand have reverberations throughout the structure of an economy and development involves as an aspect shifts in demand.

The school met this problem by assuming that innovations would be made out of surplus capital and not withdrawn from the existing circular flow. Thus J. B. Say said:

. . . it is perhaps prudent to defray the charges of experiment in industry not out of the capital engaged in the regular and approved channels of production, but out of the revenue that individuals have to dispose of at pleasure, without fear of impairing their fortune. . . .[6]

4. Senior actually talked of the "length of the period of production," but concludes that this period will be longer where "abstinence . . . is a cheap instrument of production" or, the same thing, where "profits are lower." His example, Holland, made him emphasize commercial rather than manufacturing industry but I do not feel that my inference is improper. Cf. his *Outline*, pp. 195–96.

5. E.g., Ricardo, ed. Sraffa, I, 266.

6. *Op. cit.*, p. 83. Say, by the way, suggested industry as the better sector for investment for quick returns since it ties up capital for a shorter period. While we might not accept this as such, speed of return must be taken into account in planning capital outlays.

If it were claimed that a greater speed were dictated for political reasons they would have answered, then political means must be found for absorbing the shocks. Whether political reasons for rapid industrializations are to be found of such urgency as would have convinced them of the desirability of significant interference I do not know.[7] Malthus, however, spoke of the excitation of new demand as a "plant of slow growth." In general it can be said the classical school thought of capital accumulation as a gradual process.

<div align="right">ERSKINE McKINLEY.</div>

INDIANA UNIVERSITY

7. At least in the case of England itself, in the conditions of the 1820's, it has been said that "on one thing only could most of them agree — that the pace of industrialization should be slowed down." S. G. Checkland, "The Propagation of Ricardian Economics in England," *Economica*, New Series, XVI (Feb. 1949), 44.

MARGINAL ANALYSIS, MULTI-PLANT FIRMS, AND BUSINESS PRACTICE: AN EXAMPLE*

By FRED M. WESTFIELD

I

Theoretical economists are often criticized for failing to demonstrate the practical value of their research. Economic literature is almost completely barren of expositions relating how even some of the most basic principles of economic logic find actual use in specific business situations. There exists in fact a widespread belief that the techniques and results of traditional economic theory have little or nothing to contribute to the practical solution of the businessman's concrete problems. It seems appropriate, therefore, to report an interesting example of how the general principles of marginal analysis, so fundamental in theoretical discussions of resource allocation, are actually profitably applied by men of affairs in the electric power generating industry faced with a complex allocation problem.

Typical of the modern electric power generating industry are large integrated companies that operate several geographically separated generating plants and service customers over a wide region frequently encompassing parts of several states. The plants of such a system are interconnected by a complex web of transmission lines so that electricity can be "transported" from any plant to every part of the network.

Over a twenty-four hour period the demand for electricity fluctuates sharply. Periods during which demand is less than 40 per cent of peak demand are not uncommon for even the largest systems. These demand patterns vary somewhat from day to day. In the short run the number and types of available plants are fixed, the price structure is given, and the electricity demanded must be delivered to consumers regardless of whether out-of-pocket expenses are covered by revenues. From the economist's as well as from the

operator's point of view, the important short-run decision is how to allocate the demanded output among the alternative plant capacities so as to minimize costs. The stakes are high. Even a small percentage saving in costs averaging, say, $20 per hour amounts to an annual saving of about $175,000.

Until recently losses of electricity during transmission (transport costs) were generally ignored in the solution of the problem, except in isolated cases where they were obviously great. During World War II, however, practical methods of determining the losses for routine operating purposes were developed by electrical engineers; and in 1951 the American Gas and Electric Company, a system operating over forty interconnected plants and serving customers in some 2300 communities of seven states, was pioneering in using these loss data to determine least cost production schedules for their plants. Taking explicit cognizance of these losses, equivalent to abandoning the economist's familiar "no-transport-costs" assumption, leads to estimated additional cost savings of close to $150,000 per year in the operation of this system.

Paradoxically, engineers, not economists, were called upon to analyze and solve this resource allocation problem. They worked out the correct solutions long known and understood by economists. Showing no familiarity with economic literature, they "discovered" marginal analysis after some stumbling and then proceeded to apply its principles. Their theoretical expositions lack the rigor and sophistication to which economists are accustomed; but their techniques and devices for making the discovered optimal solutions operational within the firms of the industry exhibit adroitness and expediency to which economists are unaccustomed.

We shall report, in turn, the solutions of the problem, first neglecting transmission losses, then introducing them explicitly. For each case there will be developed: (i) the economist's theoretical solution which corresponds, except for detail, to the engineer's conclusions, and (ii) the industry's methods of utilizing the theoretical results.

II

How a multi-plant firm must allocate output among plants, in the absence of transport losses and transport costs, so as to minimize costs was carefully developed some years ago in an article by Professor Patinkin and a comment thereon by Professor Leontief.[1] They con-

1. D. Patinkin, "Multi-Plant Firms, Cartels, and Competition," this *Journal*, LXI (Feb. 1947), 173–205. W. Leontief, "Multiple-Plant Firms: Comment,"

sidered a firm with n plants, each with a total cost curve of the general form

$$C_i = c_i(x_i) + F_i,$$

where C_i are the total costs of plant i, $c_i(x_i)$ its variable costs as a function of its output x_i, and F_i its fixed costs. The total costs of the firm

(1)
$$\sum_{i=1}^{n} C_i + G = C,$$

where G are fixed expenses not attributable to the operations of any one plant, must be a minimum for each level of total output demanded. This is one of those constrained minimum problems so familiar to economists.[2] The constraint is simply

(2)
$$\sum_{i=1}^{n} x_i = D,$$

D being the total output demanded.

Making the customary continuity assumptions and employing the techniques of the calculus, one can deduce the desired theoretical requirements for minimum costs. The first order conditions for a minimum tell us that if more than one of the firm's plants are in operation, each must be made to produce at a rate such that its marginal costs are equal to the marginal costs of every other plant in operation. And a plant is to stand idle only if its marginal costs when idle are higher than those of the plants in operation. These conditions must be satisfied for every level of the firm's total output.

It is useful for what follows to summarize these results by the expression

(3)
$$C_i' = \lambda \qquad (i = 1, 2, \ldots, n),$$

where C_i' stands for the marginal costs of the ith plant, and λ is a Langrangean multiplier whose value is determined by the amount of

this *Journal*, LXI (Aug. 1947), 650–51. How a monopolist should allocate his output among several separated markets in the absence of transport costs — a similar problem — was carefully worked out by T. O. Yntema, "The Influence of Dumping on Monopoly Price," *Journal of Political Economy* (Dec. 1928), XXXVI, 686–98. Almost every modern economics text might be cited for the solution of problems that have much in common with that under discussion here.

2. Cf., for example, P. A. Samuelson, *Foundations of Economic Analysis*, pp. 57 ff.

total output demanded. It is to be understood that the equality signs must be replaced with a "greater than" sign for idle plants.

In general, these conditions are not enough. The second order conditions require that (i) not more than one plant with "forward-falling" marginal costs should ever be in operation. Further, (ii) if one of the several plants in operation is producing on a falling part of its marginal cost curve, then, at these outputs, the rate of decrease of its marginal cost curve must be smaller in absolute value than the rate of increase of the horizontally summed marginal cost curves — a combined marginal cost curve — of all the other plants not standing idle.[3]

The mathematically minded reader will note that these results are readily obtained by discovering the conditions under which the Lagrangean expression

$$C - \lambda(\Sigma x_i - D)$$

subject to the constraint, takes on a relative minimum. By setting the first derivatives with respect to each x_i equal to zero one obtains the set of equations (3). Boundary minima will introduce the inequalities.

The second order conditions require that the expression

(4) $$\prod_{i=1}^{k} C_i'' \sum_{i=1}^{k} \frac{1}{C_i''} > 0 \qquad (k = 2,3, \ldots, m \leqq n),$$

where C_i'' refers to the slope of the marginal cost curve for plant i evaluated at the optimum output, be satisfied for all k, regardless how we number the m ($\leqq n$) plants in operation.[4] Part (i) of the second order conditions stated in words above is deduced from the

---

3. The next two paragraphs may be omitted without loss of continuity by those not interested in the mathematical derivation of the results.

4. The principal minors of the determinant

$$\Delta = \begin{vmatrix} C_i''\delta_{ij} & 1_i \\ 1_j & 0 \end{vmatrix}, \qquad \begin{matrix} \delta_{ij} = 0 \text{ for } i \neq j \\ = 1 \text{ for } i = j \end{matrix},$$

$1_i, 1_j =$ column and row, respectively, of ones,

as well as the determinant itself must be negative (*cf.* Samuelson, *op. cit.*, pp. 362 ff.). Evaluating Δ by the Laplace Development: $\Delta = -\sum_{i=1}^{m} A_{ii}$, where each A_{ii} is a nonvanishing cofactor of an element (i.e., of a diagonal element) in the determinant $|C_{ij}''\delta_{ij}|$, and noting that

$$A_{ii} = \frac{C_1'' C_2'' \ldots C_i'' \ldots C_m''}{C_i''}$$

tact that this expression must hold for $k = 2$; part (ii), from the fact that it must hold for $k = m$.[5]

There may be several allocations satisfying our conditions. To determine which one of these is "best," it is necessary to compare total variable costs at the various allocations and select the one with the lowest. The optimum allocations are, of course, independent of all fixed costs.
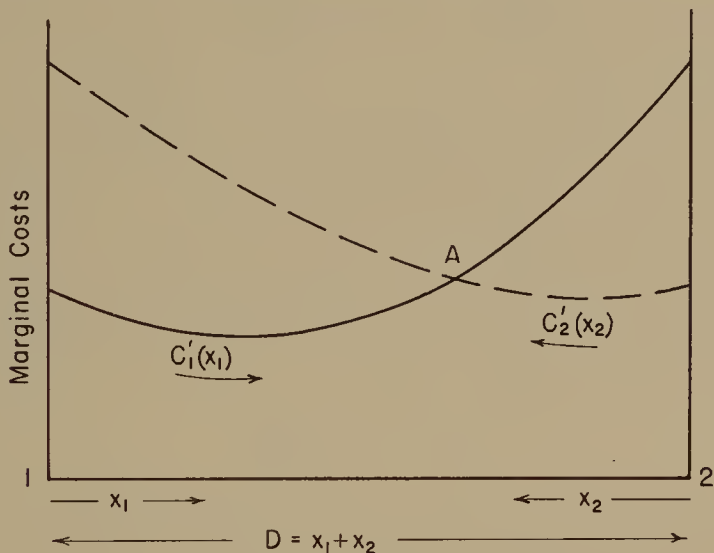

FIGURE I

The output of plant 1 (x_1) is read off from left to right, that of plant 2 (x_2) from right to left. Total output to be delivered is represented by D, equal in the absence of transport losses to the sum of x_1 and x_2. The output allocation among plants corresponding to point A, where the marginal costs of the two plants are equal ($C_1' = C_2' = \lambda$), is the least cost allocation. The combined variable costs for the two plants, represented by the sum of the areas under their respective marginal cost curves, is seen to be greater for every other allocation.

it is readily seen that the expression (4) is a statement of the conventional secondary conditions for a constrained minimum.

5. If, say, $C_1'' < 0$, then in (4) $\prod\limits_{i=1}^{m} C_i'' < 0$; but $\sum\limits_{i=1}^{k=m} \dfrac{1}{C_i''} < 0$ must then be satisfied for $k = m$, which assures that it will also be satisfied for all other (lower) k. This condition on the summation may be written $\sum\limits_{i=2}^{k=m} \dfrac{1}{C_i''} < -\dfrac{1}{C_1''}$, which

implies $\sum\limits_{2}^{m} \dfrac{1}{C_i''} < \left| \dfrac{1}{C_1''} \right|$. Hence, $\dfrac{1}{\sum\limits_{2}^{m} \dfrac{1}{C_i''}} > |C_1''|$. But the left-hand side of this

inequality is the slope of the horizontal sum of the rising marginal cost curves of the plants in operation; the right-hand side, the absolute value of the slope of the falling marginal cost curve.

If a firm has only two plants it is simple to show the theoretical conclusions on a familiar diagram (Figure I). Experimentation with different shaped marginal cost curves will convince the reader of the validity of all the conditions for minimum costs that we have stated.

III

So much for the theoretical economics. How is the problem solved in practice?

In the production of electric power, the output at each plant varies, in the short run, primarily with the energy (heat) input.[6] These relationships between the amount of heat energy provided by the fuel and the corresponding maximum amount of electric energy produced by the equipment in a plant, the production functions, are obtained from engineering data. With the technological characteristics and the price of fuel for each plant known, (short-run) marginal cost functions are readily calculated.[7] They may be written as

$$C_1' \equiv w_i \frac{1}{dx_i/dv_i} ,$$

where w_i is the price of fuel in plant i and dx_i/dv_i its marginal productivity. The units of v_i are stated in millions of B.T.U. per hour; of w_i in dollars per million B.T.U.; and of x_i in megawatts. The B.T.U. content and therefore the effective price of fuel are obtained from routine analyses.

In order to facilitate the actual job of allocating the demanded electricity or "load" among plants in conformance with our theoretical principles, a Station-Loading Sliderule was developed.[8] This mechanical device permits a centrally located dispatcher, in constant communication with the plants, to calculate quickly the optimum allocations for each demand situation.

The essentials of the sliderule, shown schematically in Figure II, are two unmovable scales, the "marginal fuel rate scale" and the "relative fuel price scale," and as many movable "plant output slides" as there are plants in the system — one for each plant. Marginal fuel rates are the additional amounts of fuel required (in millions of B.T.U.) in a plant to produce extra units of output (in megawatt-

6. In this paper we shall ignore the complications that may arise if some of the system's electricity is generated in hydro-electric plants.

7. M. J. Steinberg and T. H. Smith, *Economy Loading of Power Plants and Electric Systems* (New York, 1943), pp. 141 ff.
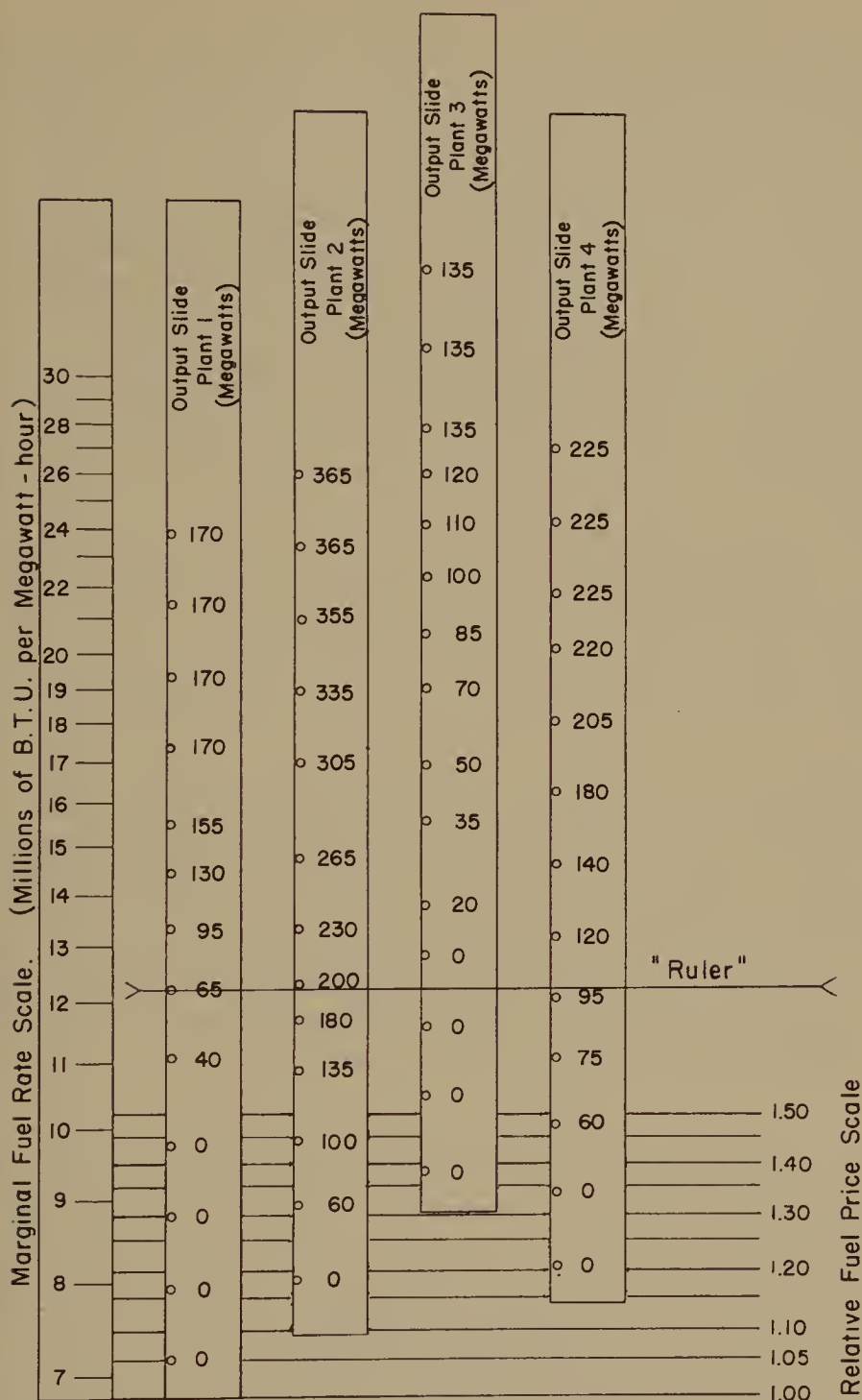
8. *Ibid.*, pp. 189 ff.

FIGURE II

The Station-Loading Sliderule.   Marginal Fuel Rates and Relative Fuel Prices are on logarithmic scales.   Plant 1 is the *numéraire* plant.   Output is demanded at the rate of 360 megawatts.   The optimum allocations to the plants are read off along the ruler.

hours). They are reciprocals of marginal productivities. Relative fuel prices are the prices of fuel at each plant relative to the price of fuel at the plant with the lowest fuel price. We may call the latter plant(s) the *numéraire* plant(s).

The marginal fuel rate scale is calibrated with the logarithms of marginal fuel rates (log dv/dx), starting at the bottom with values somewhat below those actually to be encountered and increasing continuously to some point at the top slightly greater than the largest value to be encountered in the system. Relative fuel prices are also on a logarithmic scale. This scale begins with log 1 — the index — and runs alongside the marginal fuel rate scale. Each output slide is calibrated in such a way that when the bottom of the slide is lined up with the index on the fuel price scale, its markings show the maximum output rates in the plant that the slide represents corresponding to every marginal fuel rate.

In using the sliderule the dispatcher displaces each of the plant output slides (except that of the *numéraire* plant) from the index to the appropriate point on the relative fuel price scale. Thus, if the price of fuel at the *numéraire* plant is 25¢ per million B.T.U. and at plant 2 the price is 30¢, the bottom of the output slide of plant 2 is moved to 1.2 on the relative fuel price scale. To divide the output among plants so as to "equalize" marginal cost, he merely has to move a "ruler," crossing the output slides horizontally, to that position where the sums of the outputs marked on the slides along the ruler add up to the total demand — whatever it may be at the moment. These outputs of the individual plants satisfy the first order conditions for a minimum which we discussed above.[9]

Instead of equating marginal costs of the plants in operation, the dispatcher using this device actually equates

$$(5a) \qquad \log \frac{1}{dx_1/dv_1} + \log \frac{w_1}{w_1} = \log \frac{1}{dx_2/dv_2} + \log \frac{w_2}{w_1} = \ldots .$$

But this is identically

$$(5b) \qquad \log \frac{C_1{}'}{w_1} = \log \frac{C_2{}'}{w_1} = \ldots .$$

9. This sliderule is reminiscent of the one introduced by Edgeworth (*Papers Relating to Political Economy*, II, 52–58) to illustrate Mangoldt's contribution to the comparative cost doctrine in international trade theory. *Cf.* J. Viner, *Studies in the Theory of International Trade*, pp. 458–67.

A moment's reflection will convince the reader that these equations have the same solutions as set (3).[1] Also, as inspection of Figure II will show, the required inequalities will hold for all idle plants. Using logarithms of marginal costs is convenient when there is only one variable factor (input) because the marginal cost function is simply shifted without change in shape in response to price changes of the factor.

Fortunately, the complexities introduced by decreasing marginal costs can be ignored. The very nature of plant equipment insures that marginal fuel rates and hence marginal costs are increasing functions of output. Proper use of the sliderule will satisfy the first order necessary conditions and rising marginal cost curves at every plant make these conditions also sufficient for minimum cost allocation of output.

It should be noted that no essential complications arise, either in the theoretical solution or in its application, if marginal cost curves of the plants are discontinuous.[2] In fact, in many systems some plants (the most modern and favorably located fuel-pricewise) are always used up to capacity.

IV

The foregoing solution to the allocation problem is optimal only if no variable costs or losses of electricity result from transporting the generated electricity from plants to consumers. Typically there are no variable transport costs as such once the system is built; but electrical energy is lost during the process of transmission. Such losses for a system, it was shown as recently as twelve years ago,[3] can be expressed in terms of a quadratic form

$$y = y\,(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} x_i x_j \,,$$

where y represents the loss in megawatts and the x's, as previously, are the outputs of the plants. Subsequently, methods were developed

1. Just as the equilibrium conditions in consumer theory are independent of monotonic stretchings of the "utility index," so the equilibrium conditions for a multi-plant firm are independent of such a "cost index."

2. Since, corresponding to each marginal fuel rate, the maximum output is marked on the plant's output slide, a discontinuity means that scheduled output for the plant in the neighborhood of the discontinuity will remain unchanged for several different values of the marginal fuel rate. Also, cf. Samuelson, *op. cit.*, pp. 70 ff.

3. E. E. George, "Intrasystem Transmission Losses," *Electrical Engineering* (*AIEE Transactions*), Vol. 62 (March 1943), pp. 153–58.

for deriving the loss formula coefficients, B_{ij}, by means of network analyzers — miniature representations (electrical analogues) of the entire integrated generating and transmission system.[4]

A number of simplifying assumptions are made in the development of the formula and in the estimation of the coefficients. The most significant of these, from the economist's point of view, is that during the normal course of demand variations, the demands for electricity at every point in the system are assumed to vary proportionately. This means that we may think of all the electricity actually sold to many consumers all along the system as if it were delivered to a single consumer located at a fixed delivery point. Should this assumption be unrealistic for a particular system, additional variables may be introduced into the formula, or a different set of coefficients may have to be used for each pattern of demand variations. But these complications will not concern us in this paper.[5]

The essence of the theoretical solution which makes allowances for transmission losses was worked out quite recently by electrical engineers.[6] Again the problem is to minimize the firm's total costs, represented by equation (1), for each level of demanded output. But the simple "conservation-of-product constraint" (equation (2)) is no longer appropriate. There are losses which depend upon how much electricity each plant is producing. The new constraint becomes

$$(2') \qquad \sum_{i=1}^{n} x_i = D + y(x_1, x_2 \ldots x_n).$$

It simply says that the total electricity generated by all the plants together must add up to the total quantity demanded by consumers plus the quantity lost in transmission.

Proceeding as before, we can readily discover the necessary and sufficient conditions for minimum cost allocation.[7] The first order

4. J. B. Ward, J. R. Eaton, H. W. Hale, "Total and Incremental Losses in Power Transmission Networks," *AIEE Transactions*, Vol. 69 (Part I, 1950), pp. 626–31.

5. *Ibid.*, p. 629.

6. E. E. George, H. W. Page, and J. B. Ward, "Coordination of Fuel Cost and Transmission Loss by Use of the Network Analyzer to Determine Plant Loading Schedules," *AIEE Transactions*, Vol. 68 (Part II, 1949), pp. 1152–60. L. K. Kirchmayer and G. W. Stagg, "Evaluation of Methods of Co-ordinating Incremental Fuel Costs and Incremental Transmission Losses," *AIEE Transactions*, Vol. 71 (Part III), 1952, pp. 513–20.

7. The Lagrangean expression to be minimized under constraint is
$$C - \mu(\Sigma x_i - y - D).$$

conditions can be summarized by

(7a) $$C_i' - \mu(1 - y_i) = 0,$$

where y_i is the marginal transmission loss from plant i $(= \partial y/\partial x_i)$; it is the percentage of an extra unit of power produced at plant i that is lost during transmission if the output of all other plants remains the same. By rewriting these relations in the form analogous to (3)

(7b) $$\frac{C_i'}{(1 - y_i)} = \mu,$$

where it is again understood that the equal sign must be replaced with a "greater than" sign if the ith plant is idle, we see, since $1 - y_i$ represents the marginal percentage of the power produced in plant i delivered to the consumer, that the new Lagrangean Multiplier μ is the marginal delivered cost of electricity — the marginal c.i.f. cost, so to speak.

It is now simple to state these results in words. If more than one plant is being operated, whatever the demanded output may be, each must be operated at a rate such that the marginal delivered cost of its output (the marginal cost of the output generated by it actually reaching the consumer) is equal to the marginal delivered cost of the output of each and every other plant in operation. A plant should be kept idle only if the marginal delivered cost of its output when idle is greater than the marginal delivered cost of those actually operating.

The necessary and sufficient second order conditions are not easily stated in words.[8] However, if we know that all the marginal production cost curves are rising functions of output, the marginal conditions above are sufficient to assure us of a regular minimum solution. As we have indicated, the marginal production cost curves of electric generating plants are by their very nature this type of function.

Marginal transmission losses can also be expressed as marginal transmission (transport) costs. Rewriting equations (7b) in the equivalent form

8. Letting $\partial^2 y/\partial x_i \partial x_j = y_{ij}$ and using the notation of note 4, p. 256, the necessary and sufficient conditions for the minimum are that the bordered determinant

$$\begin{vmatrix} C_i''\delta_{ij} + \mu y_{ij} & 1_i - y_i \\ 1_j - y_j & 0 \end{vmatrix}$$

and all its principal minors be less than zero. As the quadratic form y (the loss formula) is intrinsically positive, it is seen that these conditions do hold if all C_i'' are non-negative.

(7c)
$$C_i' + \mu y_i = \mu,$$

the marginal delivered cost, μ, is seen to be the sum of the marginal production cost, C_i', and the marginal transmission loss, μy_i — the marginal transmission loss "priced" at the marginal delivered cost.

Before proceeding with the explanation of how these results are brought to bear in actual practice in scheduling production of the generating plants, it may be helpful to show our theoretical solution graphically for a two-plant firm. For this purpose a diagram different from that used to illustrate the no-transmission-loss-case is convenient. In Figure III the electricity produced by plant 1 is plotted on the horizontal axis, that produced by plant 2 on the vertical. The family of curves labeled C^a, C^b, C^c, etc., are what I call Constant Outlay Curves. For a given rate of total expenditure by the firm, say C^b, the curve labeled C^b shows the maximum rate of output that
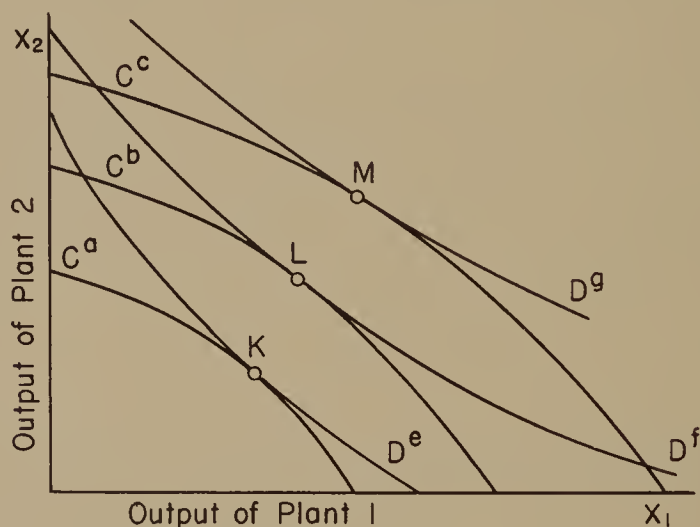
FIGURE III

The points K, L, M, represent least-cost allocations of demanded output rates D^e, D^f, D^g, respectively. No allocations can be found for these demand situations which lie on lower Constant Outlay Curves.

can be produced in plant 2 for each specified output rate of plant 1. The further such curves are away from the origin, the greater is the outlay (total expenditure) identified with them. Their slopes are $-C_1'/C_2'$ — the ratio of marginal production costs in plant 1 to those in plant 2; and if we know that marginal production costs in the plants are increasing functions of output, we can be sure that the Constant Outlay Curves are concave to the origin. The other set of curves, D^e, D^f, D^g, etc., are what we may call Allocation Possibility Curves. If consumers demand electricity at the rate, say D^e, the

D^e curve represents all the possible ways of dividing total production of the firm (greater because of losses than the rate of demand) among plants such that in fact the rate D^e reaches the consumers. The greater the consumer demand, the further away from the origin is the relevant Allocation Possibility Curve. If there are transmission losses, all of these curves must be convex to the origin. Their slopes are $-(1 - y_1)/(1 - y_2)$ — the ratio at the margin of the percentage of electricity produced in plant 1 reaching the consumer to the percentage of electricity produced in plant 2 reaching the consumer. If transmission losses are neglected, $y_1 = y_2 = 0$ and the Allocation Possibility Curves are negatively sloped 45° lines.

The optimum allocations are all those points on the graph where the Allocation Possibility Curves touch but do not cross the Constant Outlay Curves. All other allocations for specified demand situations represented by other points on the Allocation Possibility Curves will lie on higher Constant Outlay Curves; hence such allocations cost more. It is seen that the optimum points are precisely the ones specified by the previously stated marginal conditions (equations (7b)).

V

Attempts to put the theoretical results of this more complex problem into practical operation without modifications would be fraught with difficulties. There are $n + 1$ nonlinear relations (i.e., equations (2') and (7c)) which would have to be solved simultaneously for the n plant outputs and for μ, for each demand situation. The costs of such computations primarily because of the nonlinearities are likely to be very large.

In actual practice marginal production cost curves of generating plants can be adequately approximated up to maximum output by straight lines

$$C_i' = a_i + b_i x_i, \qquad a_i, b_i > 0.$$

Nordin[9] showed this by a careful statistical study and engineers[1] have independently come to the same conclusion. But substitution of these linear relations into equations (7c) will show that this good fortune does not eliminate the difficulties. There are still $n + 1$ equations, and they are still nonlinear.

Two approximate solutions have been suggested by the engineers.

9. "Note on a Light Plant's Cost Curves," *Econometrica* (July 1947), Vol. 15, pp. 231–35. The findings imply that a rising linear function of output gives significantly better estimates of marginal costs in the plant that Nordin studied than polynomials or a constant.

1. George, Page, and Ward, *op. cit.*, p. 1153. See also the discussion of this paper, pp. 1160–61.

Both have been verified by actual sample calculations to yield alloca-
tions extremely close to the theoretical optimum. The first, a linear
approximation,[2] is obtained by replacing on the left of each equation
(7c) the variable μ with a constant γ:

(8a) $$C_i' + \gamma y_i = \mu.$$

Making use of the linear estimates of the marginal cost curves and
substituting for y_i the partial derivatives of the loss formula this
becomes

(8b) $$a_i + b_i x_i + \gamma(2 \sum_{j=1}^{n} B_{ij} x_j) = \mu.$$

The new constant γ is a weighted average value of marginal delivered
costs μ and is used to "price" the transmission losses. Note that these
are linear relations in the unknowns. Even if inequalities replace
some equalities for certain values of demanded output, the allocations
corresponding to each μ, hence by equation (2′) for each level of
delivered output, may be obtained without difficulties by the use of
modern high-speed electronic computing equipment.

The other approximation, now actually used in the American
Gas and Electric Company System, is called the "Penalty Factor
Method."[3] Rather than "pricing" transmission losses at actual mar-
ginal delivered costs (μ), as in the exact solution, or at some average
value of marginal delivered costs (γ), as in the linear approximation,
this method charges for them at the marginal production costs of each
plant in question (C_i'). The equations (7c) then become

(9a) $$C_i' + C_i' y_i = C_i'(1 + y_i) = \mu;$$

or substituting for y_i,

(9b) $$C_i'(1 + 2 \sum_{j=1}^{n} B_{ij} x_j) = \mu.$$

The term $1 + y_i$ is called the "penalty factor" for plant i.[4] These
penalty factors can be efficiently calculated for each plant under
numerous operating conditions of the system.[5]

An advantage of this method is that it permits continued use of
the Station-Loading Sliderule, which, as will be recalled, does not
require linear approximations or, for that matter, continuity of the

2. *Loc. cit.*
3. L. K. Kirchmayer and G. H. McDaniel, "Transmission Losses and
Economic Loading of Power Systems," *General Electric Review,* Vol. 54 (Oct.
1951), pp. 39–46. Also, Kirchmayer and Stagg, *op. cit.*
4. The "penalty factor" consists of the first two terms of the expansion
of $1/1 - y$ in the exact formulation, equations (7b).
5. *The Wall Street Journal,* August 12, 1953, p. 18, reported the purchase
of a \$110,000 electronic computer by the American Gas and Electric Company
for calculating "penalty factors."

marginal production cost curves. Taking logarithms of the set of equations (9a) and writing it in the form of equations (5a):

$$(10) \qquad \log \frac{1}{dx_1/dv_1} + \log \frac{w_1}{w_1} + \log (1 + y_1)$$

$$= \log \frac{1}{dx_2/dv_2} + \log \frac{w_2}{w_1} + \log (1 + y_2) = \dots,$$

one sees that the penalty factors, once known, play a role much like relative fuel prices and can be handled by the dispatcher in a similar manner.

He proceeds by first obtaining a "trial allocation schedule" which neglects transmission losses in precisely the manner described earlier. The output slide for each plant is then moved upward on the relative fuel price scale by the value of its penalty factor corresponding to this trial allocation. A new allocation schedule is read off and each output slide is now displaced by the change in its penalty factor resulting from the change in allocations. This cycle is repeated until for every plant the penalty factors used on the fuel price scale correspond to those penalty factors called for by the allocation schedule obtained with their use. This iterative process converges rapidly because penalty factors are generally small compared to relative fuel prices. Moreover, in determining allocations on an hour-to-hour basis only comparatively small changes in total demand are usually involved. By starting with the allocation schedule of the previous period rather than with the trial schedule obtained by neglecting transmission losses entirely, the changes in penalty factors are even smaller.[6]

While the allocations obtained by either of the two approximate methods do not vary for practical purposes from those obtained by solving the exact nonlinear equations, they may differ markedly from those obtained when transmission losses are neglected. Results of sample calculations, reproduced in Figure IV, for a simplified eight-plant representation of the American Gas and Electric Company's network of almost fifty plants show the estimated additional savings to the firm under the various demand conditions when in determining allocations cognizance is taken of transmission losses. It is estimated that savings in production costs, or, perhaps more appropriately, in

6. It is not clear to the author why the approximate $(1 + y_i)$ is calculated as the penalty factor rather than the exact $(1/1 - y_i)$. The latter could also be used on the sliderule. Its use would involve no mathematical approximation and should be no more difficult to calculate. Correspondence with the American Gas and Electric Service Company indicates that this system is now (while this article is in proof) using this exact penalty factor method.

transportation costs will average about $150,000 per year. These benefits are apparently sufficiently large to justify the costs of the additional computations.
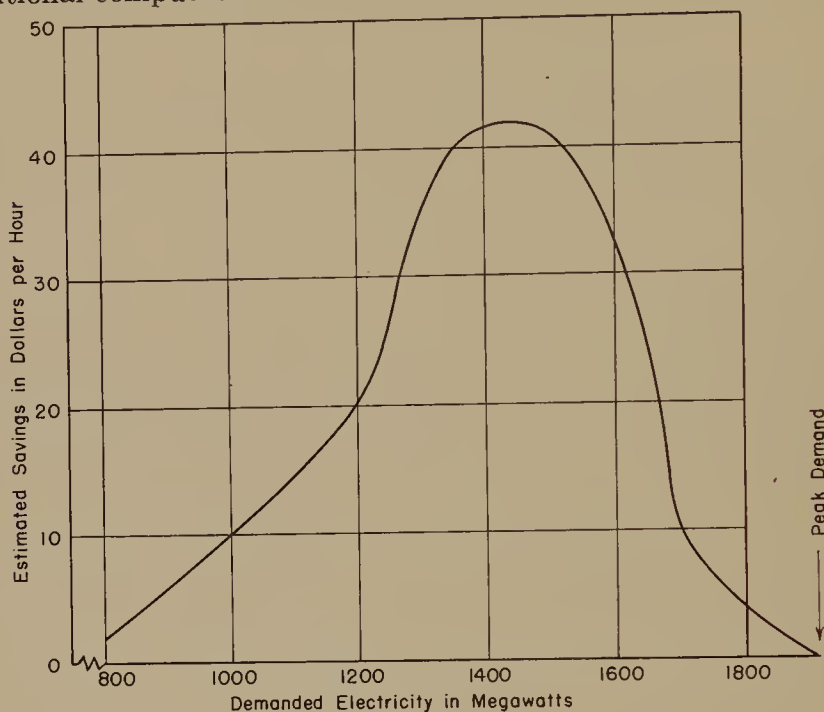

FIGURE IV

Cost reductions from taking account of transmission losses in model of American Gas and Electric Company System.
Source: Kirchmayer and Stagg, *op. cit.*, p. 519.

VI

Fictitious illustrations of the theoretical constructs of classical economics abound; authentic applications and tests are rare. In these pages we have come to grips with a specific problem in marginal analysis that has a very real counterpart in the day-to-day management of firms in the electric power industry.

We proceeded within the framework of the classical model of the profit motivated firm. Using the hypothesis of cost minimization we were able to specify how such a model firm would apportion production among the several plants under its control. Two cases were distinguished. For each, the model by any reasonable standards proved itself not only as a predictive device, but also showed itself of fundamental importance for solving the problem in practice.

FRED M. WESTFIELD.

NORTHWESTERN UNIVERSITY

THE REGIONAL IMPACT OF MONETARY POLICY[*]

By IRA O. SCOTT, JR.

The shortcomings of aggregative analysis should be recognized in the field of monetary policy as in other areas of economics. Indeed, if open market operations are compared with, say, budget policy, the limitations imposed by aggregation are relatively more severe because of the important roles played by the Federal Reserve Bank of New York and the New York money market. In this paper an attempt will be made to discover whether significant lags exist in the transmission of the effects of credit policy from New York to the rest of the country. First, the time-pattern of free reserves will be examined on a district and bank-class basis. Next, a similar analysis will be made of commercial bank portfolio policy. Finally, possible interpretations of these empirical findings will be discussed along with their corresponding policy implications.

The time-scope of the study has been limited to the period June 1951 through May 1953. The former date was chosen because it marks approximately the transition to a more flexible credit policy following the completion of support operations connected with the Treasury-Federal Reserve Accord.

I. CHANGES IN FREE RESERVES

Evidence of lags in the distribution of the effects of monetary policy will be sought first in data showing the time-pattern of free reserves for various groupings of member banks. It will be assumed that a policy of credit restraint tends to have its initial effect upon the Central Reserve City banks.[1] The pattern of gains and losses to free reserves in this segment of the banking system may thus be used as a base from which time lags of other groupings may be determined.

[1] This assumption seems to be justified in the first place by the proximity of these banks to System open market operations. Secondly, this category of banks exhibits a peculiar sensitivity to a change in credit tightness at any point in the banking system. This sensitivity, of course, stems from the correspondent relationship.

It is also assumed that developments in New York City and their lagged permeation of the whole economy are indicative of System policy, but no attempt is made to determine precisely the nature of policy at any specific time.[2]

In this portion of the study, the magnitude of free reserves is used as an inverse index of the effectiveness of a restrictive policy. Free reserves are defined as total reserves less required reserves less borrowing from Federal Reserve Banks. Presumably, the traditional reluctance of member banks to remain in debt to the System could be relied upon to prevent an increase in borrowing relative to excess reserves (total less required reserves) if other means of keeping total reserves in line with required reserves were readily available. An increase in borrowing relative to excess reserves — that is, a decrease in free reserves — may then be taken as evidence of more stringent credit conditions resulting from increased difficulty in disposing of marketable assets to provide necessary reserves.[3]

Semi-monthly averages of daily figures for free reserves have been charted for various member bank groupings (Charts I–IV). Total free reserves of the member banking system (Chart I*A*) fluctuated around a fairly level positive trend until April 1952, when a downward movement appeared. After a lower turning point in December 1952, they followed a slightly rising trend, but remained on a negative level, during the rest of the period.

In contrast with the aggregative pattern, free reserves of Central Reserve City banks (Chart I*B*) fluctuated around the zero level. Excepting the momentary recovery of March 1952, these banks sustained increasing pressure, which became extremely great in December 1952 and January 1953, when a reversal in the trend of free reserves appeared. Free reserves were particularly low for this group of banks from November to mid-April 1952–53.

Data for Reserve City banks (Chart I*C*) reveal a pattern similar to that of the Central Reserve category, with a somewhat milder reversal in trend during the latter part of the period.

Contrary to Central Reserve and Reserve City patterns, Country banks (Chart I*D*) had positive free reserves during the entire period. Moreover, the latter's free reserves appeared to follow a

---

2. The nature of System policy at a particular point in time may be difficult to ascertain, since it may or may not involve overt action.

3. There may, however, be some tendency toward seasonal shifts in the member banks' borrowing behavior parameter. Thus, banks may borrow more freely in the second half in anticipation of a seasonal relief of reserve pressure in the first half of the following year.
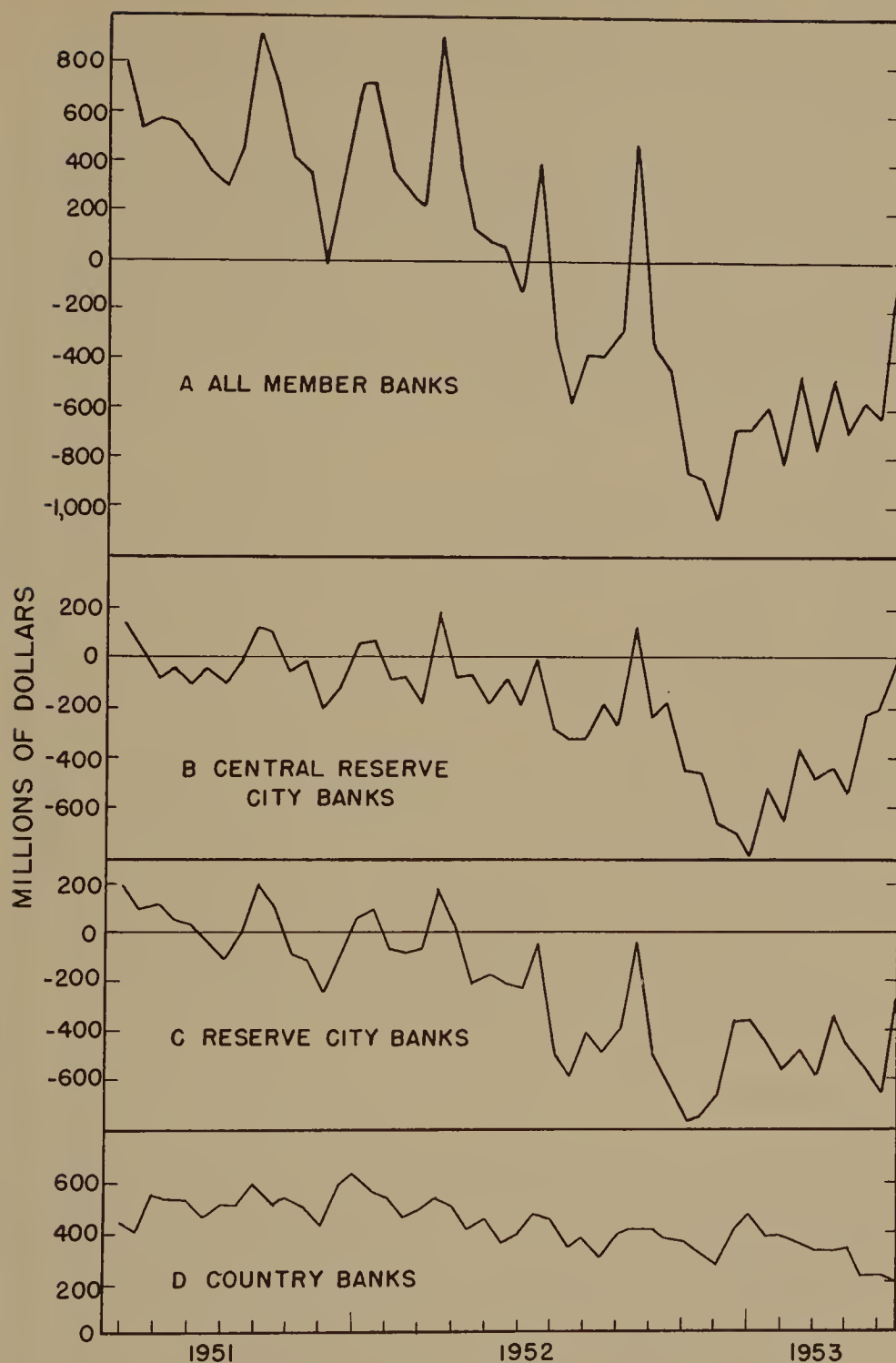
CHART I

Free Reserves of All Member Banks and by Bank Class[1]

1. Semi-monthly averages of daily figures, June 1951–May 1953, in millions
of dollars.   Free reserves are defined as excess reserves less borrowing from Fed-
eral Reserve Banks.   Source: Board of Governers, Federal Reserve System.

rising trend during the last half of 1951, and then, in January 1952, began a gradual decline that continued throughout the remainder of the period.

The free reserve patterns for these three classes of member banks reveal a significant lag in the Country bank component as compared with the Central Reserve and Reserve City groups. The Country bank pattern shows a slightly rising trend, while the other two categories have level or declining trends in the earlier part of the period. Whereas a lower turning point is reached in the Reserve City and Central Reserve City data by the end of 1952 or early 1953, there is
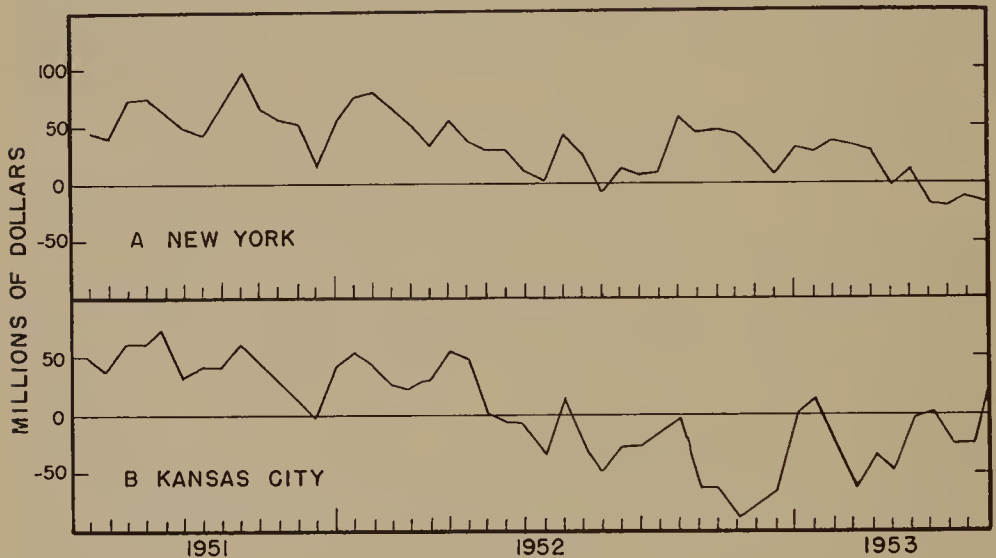
CHART II

FREE RESERVES OF RESERVE CITY AND COUNTRY BANKS,
SELECTED DISTRICTS[1]

1. See note to Chart I.

no manifestation of such a trend reversal in the latest Country bank data.

Free reserves of member banks are next charted by Reserve district (Charts II–IV). In order to eliminate the influence of the money-market banks from the New York and Chicago District totals, Central Reserve City banks were excluded. Reserve City plus Country bank data for the New York and Chicago Districts could then be compared with the Central Reserve City pattern.

It will be recalled that the trend of free reserves in Central Reserve City banks declined gradually through October 1952, then precipitously in the latter part of 1952 and early 1953, with consider-
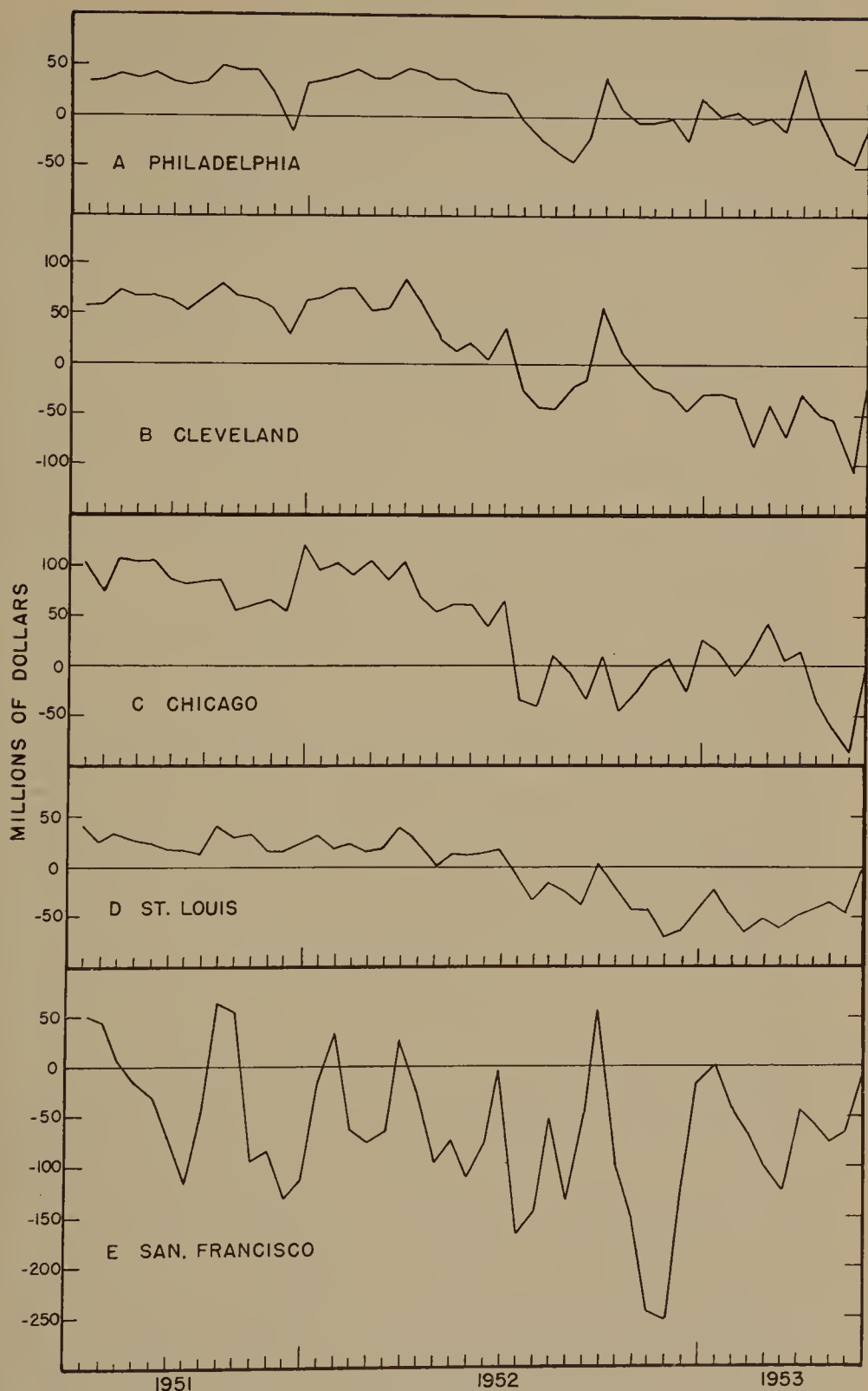
CHART III

FREE RESERVES OF RESERVE CITY AND COUNTRY BANKS,
SELECTED DISTRICTS[1]

1. See note to Chart I.

able gains in the closing months of the period being surveyed. With the exception of a few periods, at discrete intervals, free reserves were negative for this class of member banks.
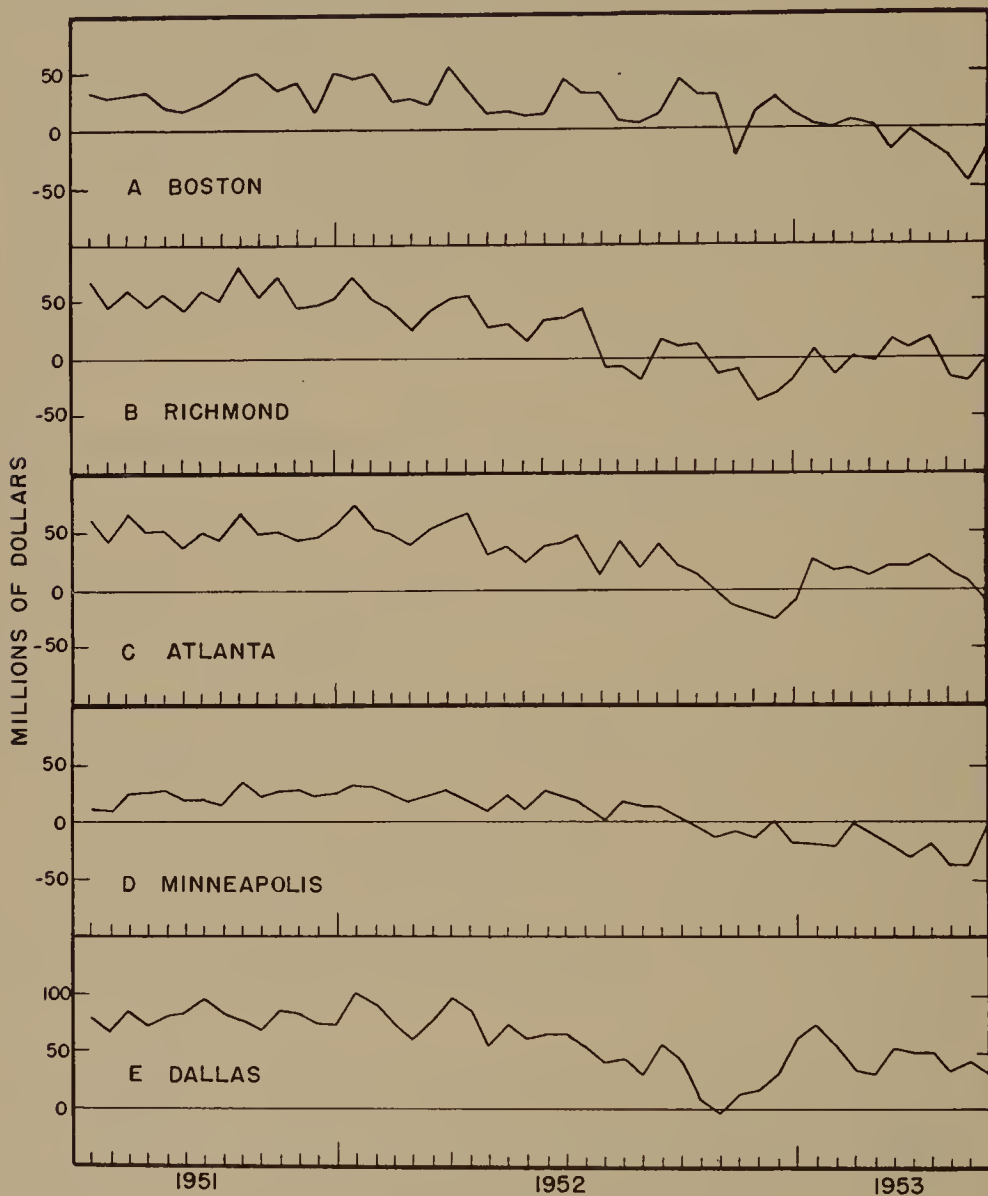

CHART IV

FREE RESERVES OF RESERVE CITY AND COUNTRY BANKS,
SELECTED DISTRICTS[1]

1. See note to Chart I.

The other groups of member banks differed from the Central Reserve City pattern in two respects. First, the former displayed less erratic fluctuations in free reserves. Secondly, the magnitude of

free reserves was predominantly positive. The San Francisco District (Chart III*E*) was the one exception, being generally negative, with a wide range of fluctuation.[4]

With respect to trend, roughly three groups of districts may be distinguished. First, there seemed to be a slightly negative trend in the New York (Reserve City and Country, Chart II*A*) and Kansas

TABLE I

INITIAL MONTH DURING WHICH FREE RESERVES
BECAME ZERO OR LESS

July 1951	Total Central Reserve City	(Chart I*B*)
	San Francisco	(Chart III*E*)
December 1951	Philadelphia	(Chart III*A*)
	Kansas City	(Chart II*B*)
April 1952	St. Louis	(Chart III*D*)
July 1952	New York (Reserve City and Country)	(Chart II*A*)
	Cleveland	(Chart III*B*)
	Richmond	(Chart IV*B*)
	Chicago (Reserve City and Country)	(Chart III*C*)
October 1952	Atlanta	(Chart IV*C*)
	Minneapolis	(Chart IV*D*)
	Dallas	(Chart IV*E*)
November 1952	Boston	(Chart IV*A*)

City (Chart II*B*) Districts, with some evidence of recovery at the end of the period in the latter district. Kansas City, therefore, was similar to the Central Reserve City pattern with respect to trend.

The second group displayed a relatively level trend during approximately the first half, but increased pressure in the latter half of the period. This group is composed of the following districts: Philadelphia (Chart III*A*), Cleveland (Chart III*B*), Chicago (Reserve City and Country, Chart III*C*), St. Louis (Chart III*D*), and San Francisco (Chart III*E*).

The third group consists of Boston (Chart IV*A*), Richmond (Chart IV*B*), Atlanta (Chart IV*C*), Minneapolis (Chart IV*D*), and Dallas (Chart IV*E*). The distinguishing feature of this group was the appearance of a slight easing tendency during the earlier part of the period. Thereafter, more or less increased pressure was apparent.

The movement of free reserves into the negative range marked one stage in the ever-tightening band of pressure felt by the member banks as the Federal Reserve System pursued a hard money policy.

4. The similarity between the San Francisco District and Central Reserve City patterns may possibly be attributed to the predominance of the Bank of America in the former group.

A tabulation has been made (Table I) showing the initial date on which free reserves fell to or below zero. A cursory inspection of the table reveals considerable lags behind the Central Reserve City category in all districts except San Francisco.

In recapitulation, the trend of the first group (Chart II), consisting of two districts, resembled rather closely the Central Reserve City trend. The second group, composed of five districts (Chart III), showed some lag in downward trend behind the Central Reserve City pattern. The remaining districts, comprising the third group (Chart IV), manifested even greater insensitivity to the stringencies of the money market in displaying an absence of pressure on free reserves in the earlier half of the period. At no time during the period being studied did any of the districts, with the exception of San Francisco, sustain the same degree of pressure as the Central Reserve City group. Finally, the appearance of a zero or negative free reserve position lagged behind the Central Reserve City group in all Districts, again with the exception of San Francisco.

II. Commercial Bank Portfolio Policy

According to the description of commercial bank behavior at the beginning of the previous section, a commercial bank will tend to increase its indebtedness to the System when confronted by urgent customer loan requests combined with a condition of stringency in the money market. The redistribution of assets implied by these varying pressures will undoubtedly involve some decline in investments relative to loans. Assuming that such a portfolio policy is generally characteristic of commercial banks, evidence of lags in the regional distribution of the effects of a restrictive policy may be sought in the changing pattern of portfolio composition in the various Federal Reserve districts.

Charts have been constructed to depict the pattern of selected assets of various categories of weekly reporting member banks on first and third Wednesdays during the period under review[5]. A comparison of regional charts with one showing totals of these various assets for all districts revealed a great diversity of portfolio patterns, particularly with respect to the relative positions of loans and investments.

As suggested earlier, the weekly reporting member banks in New York City are presumably among the first to react to the Sys-

5. These charts are omitted here because it seemed possible to summarize them in the text and thus avoid imposing an undue burden upon the reader.

tem's credit policy moves. Hence, the pattern of selected assets for these banks was used as a standard. At the beginning of the period, net loans in New York City banks were at roughly the same level as investments. A growing divergence between these two asset categories occurred during the second half of 1951. Subsequently, the level of net loans remained at about the same magnitude above total investments until September 1952. During the remainder of the period, the difference continued to grow. Until about September 1952 total investments followed a fairly level trend, declining thereafter until the end of the period. A gradual upward trend in net loans was apparent throughout. The pressure on investments thus appears to have been relative through August 1952. Thereafter, it became absolute.

To a great extent, changes in business (commercial, industrial, and agricultural) loans and Government securities accounted for these movements of net loans and investments, respectively. At the beginning of the period, business loans were substantially less than holdings of Governments. The gap was closed by October 1951, with the two fluctuating at roughly the same level through August 1952. Afterwards, Governments declined, while business loans expanded until near the end of the period. Again, the pressure on Governments appeared to be largely relative through August 1952, becoming absolute from that point on until the latter part of the period.

The composition of the portfolios of the remaining groups of weekly reporting member banks could then be compared with this pattern. Philadelphia was the first district to be compared with the New York City pattern. As in the latter case, a change in portfolio policy occurred around August 1952 with absolute pressure upon investments thereafter. Prior to this change in policy, and contrary to the basic pattern, no relative pressure upon investments was apparent. With the exception of a single reading at the beginning of the period, net loans exceeded investments in New York City, whereas they did not reach this level until October 1952 in the Philadelphia District. Moreover, Governments were continuously greater than business loans in the latter area. It will be recalled that this relationship was reversed after August 1952 in New York City.

In the New York District (less New York City) the pattern was similar to that of Philadelphia, where no relative pressure appeared prior to a change in portfolio policy in the neighborhood of August 1952. Afterwards, however, only a relative pressure upon investments was apparent, as net loans expanded more rapidly than invest-

ments.   During the entire period, investments and Governments exceeded net loans and business loans, respectively.

The Boston and Richmond Districts provided the third portfolio pattern.   They both showed a change in policy around August 1952. Afterwards, investments followed a declining trend.   In these two respects, the Boston and Richmond patterns resembled those of New York City and Philadelphia.   Earlier, and contrary to the New York District, as well as to New York City and Philadelphia, Boston and Richmond had relative increases in their holdings of investments. Net loans were less than investments throughout in the case of Richmond, and they did not overtake investments until January 1953 in the Boston District.   Business loans were less than Governments during the entire period in both districts.

The Atlanta, St. Louis and San Francisco Districts comprised the next category that was considered.   Like the previous group examined, these districts manifested some relative gain in investments prior to a change in portfolio policy in the neighborhood of August 1952.   This was followed immediately, however, by only a relative pressure upon the investment portfolio.   A declining trend in the latter did not appear in Atlanta and San Francisco until about November 1952, and not until January 1953 in St. Louis.   Net loans were generally greater than investments in the St. Louis and San Francisco Districts.   In Atlanta, on the other hand, this relationship was reversed. Governments were greater than business loans throughout the period in all three districts.

The next group could be distinguished from earlier patterns in that the apparent change in portfolio policy occurred at a later time — February 1953.   Before this date, both Minneapolis and Kansas City had relative gains in investments; afterward, absolute declines. In the Minneapolis District, investments were greater than net loans except for brief intervals at the beginning and at the end of the period. Otherwise, investments exceeded net loans and Governments were greater than business loans throughout the period in both districts.

This left the Cleveland, Chicago (less city of Chicago), and Dallas Districts as not elsewhere classified.   The portfolio policies followed by the member banks in these districts revealed even less relationship to the standard pattern.   In the Cleveland District, a gradual expansion of net loans characterized the whole period.   A downward trend in investments beginning in January 1952 was followed by relative gains toward the end of the year and then a second decline beginning in January 1953.   At no time did net loans exceed investments or

business loans exceed Governments. Essentially the same description, with some differences in timing, seemed appropriate in the cases of Chicago and Dallas.

A tabulation has been made (Table II) of the portfolio response reflected in the degree of pressure exerted by net loans upon investments in all districts except those in the not-elsewhere-classified category. In summary, these charts of selected assets of the weekly reporting member banks revealed the emergence of a change in portfolio policy in a majority of districts during late Summer or early Fall 1952. However, when changes in the degree of pressure upon the investment portfolio were taken into account, significant lags were evident. Remaining districts either showed a later change in portfolio policy or practically no relationship at all to the New York City pattern. Moreover, with the possible exceptions of Dallas and San Francisco, the outlying districts were always subject to a smaller degree of pressure, as measured by the relationship between loans and investments, than were the New York City banks.[6]

TABLE II

PORTFOLIO RESPONSE REFLECTED IN THE DEGREE OF PRESSURE UPON INVESTMENTS

	Before Change in Portfolio Policy	Date of Change in Portfolio Policy	After Change in Portfolio Policy
New York City	Relative pressure	August 1952	Absolute pressure
Philadelphia	No relative pressure	August 1952	Absolute pressure
New York (less New York City)	No relative pressure	August 1952	Relative pressure
Boston Richmond	Relative increase in investments	August 1952	Absolute pressure
Atlanta St. Louis San Francisco	Relative increase in investments	August 1952	Relative pressure
Minneapolis Kansas City	Relative increase in investments	February 1953	Absolute pressure

In order to examine the portfolio policy of commercial banks outside the weekly reporting classification, last Wednesday figures for United States Government securities and loans of weekly report-

6. The use of such a criterion is, of course, fraught with difficulties. For example, one would expect to find a greater proportion of Governments in longer maturities where small banks predominate. In such cases, greater pressure would be required to force the sale of Government securities.

ing member banks were substracted from corresponding figures for all commercial banks. Holdings of Government securities were greater than loans throughout the period, though the trend of the difference seemed to be slightly downward. The third quarter of 1952 squeeze which was evident in many of the weekly reporting bank data did not appear until February 1953, if a seasonal tightness in December 1952 is excepted. Thus, the data showed a considerable lag as compared with the weekly reporting member banks, revealing a more tenuous connection between the nonreporting commercial banks and the money market.

III. Policy Implications

Any conclusions based upon an analysis of the data presented in previous sections of this paper must be qualified because of the limited amount of data available on a regional basis. For example, the isolation of credit policy effects would require a knowledge of other factors, such as interregional trade flows, which influence local credit conditions. However, if the various assumptions made during the foregoing analysis are deemed acceptable, significant lags in the regional impact of monetary policy appear to have occurred. Whether the pattern of free reserves or the loan-investment relationship is used as a guide-post, a condition of credit tightness at the center had a delayed effect upon the outlying districts with few exceptions.

A preference for monetary policy is frequently expressed in the literature on the ground that it is a more flexible instrument of control than budget policy. The central bank can maintain daily, even hourly, contact with the money market, and policy adjustments can be effected as quickly. Tax and expenditure policy, on the other hand, is not amenable to such facile manipulation. However, lags in the transmission of money-market stringencies to the rest of the economy may require a reappraisal of the relative effectiveness of monetary policy. To some extent, these lags may underlie the scepticism with which many view monetary policy as an instrument of control.[7]

What, then, are the policy implications of these findings? Before considering various policy prescriptions, the different ways in which

7. Economists sometimes criticize writers in the financial press for what the former deem to be a hypersensitivity to monetary policy. The latter often find the economists' position to be untenable. Perhaps both are right. Someone does have to face the "whirlwind" in New York. It does not follow, however, that there is a noticeable effect upon current investment activity in, say, Houston, Texas.

the effect of, say, open market sales may be transmitted, will be discussed. Four cases may be distinguished.

(1) The System Open Market Account sells directly to the New York agents of purchasers in all of the districts. Payment is made with federal funds obtained from the Federal Reserve Bank in each district. There is no lag, since member bank reserves in all districts fall at the same time.

(2) The effect of a change in monetary policy is transmitted outward, in centrifugal fashion, from the money-market center. As the System Open Market Account sells, dealers and others at the center buy. A money-market stringency develops. Eventually, funds are attracted to the center via the System's Inter-district Settlement Fund as New York dealers move their inventories to out-of-town investors. The effect of the stringency is passed along in a lagged manner to the other Federal Reserve districts. The center may continue to sustain a relatively greater degree of pressure if further open market sales are made.

(3) Sales are made to widely dispersed investors, who obtain federal funds through their New York correspondents. A credit stringency emerges at the center as pressures from outlying regions converge, in centripetal fashion, upon the New York money market. Whereas the contraction of bankers' and other correspondent balances imposed upon a particular district is commensurate with that district's participation in the open market transactions, the money-market banks bear the brunt of the whole operation. The length of the lag in the subsequent decline of non-money-market member bank reserves depends upon the rapidity with which correspondent balances are replenished.

(4) Dealers at the center expand their positions as the System reduces its holdings. Reserves of the money-market banks are down accordingly. Eventually, dealer inventories decline as investors throughout the economy increase their holdings by drawing on their accounts with correspondent banks. A further delay in the transmission process ensues before these accounts are restored to previous levels.

In cases (1) and (3), sales are made immediately to widely dispersed buyers. New York dealers are the initial investors in (2) and (4). (1) and (2) thus resemble (3) and (4), respectively, except for the use of the correspondent banking system in the latter two cases.

No lag appears in case (1). In (2) the lag is due to the time involved in the percolating process as the securities filter through

the New York money market. The lag of case (3) relies exclusively upon the existence of a correspondent relationship and the delay in rebuilding correspondent balances. The lag in case (4) is compounded of both the percolating effect of (2) and the correspondent banking effect of (3).

The data clearly show case (1) not to be dominant. It may also be inferred from the portfolio figures that the lags cannot be attributed solely to the correspondent ralationship. That is, in making investment decisions, a decline in bankers' balances would rank in importance with a decline in legal reserves. Thus, the lag in portfolio policy points to the significance of the percolating effect. Except for these general comments, it is impossible, on the basis of available information, to assess more accurately the relative importance of these cases.

The problem appears to be chiefly one of compensating for obstacles which prevent a rapid centrifugal movement of money market expansions or contractions. The effect of open market operations could be dispersed more quickly if capital markets could be developed in, say, each Federal Reserve district. Sales and purchases of securities could then be allocated to and effected in the various districts simultaneously.

To the extent that the correspondent system is the seat of the difficulty, there is less need for developing regional capital markets. The latter, however, would involve the growth of correspondent relationships on a regional basis. This would eliminate some of the extremes in pressure to which the New York money market is now subjected, though the necessity for a correspondent relationship with New York founded upon the needs of trade and commerce would remain.

It is significant that dealers now hold positions in New York and that they encounter considerable difficulty in borrowing funds from outlying banks. If regional capital markets existed, with positions held and financed locally, a greater degree of perfection could be achieved in the movement of funds.

Such a change in the execution of open market policy would presumably entail the introduction of greater flexibility in the imposition of reserve requirements upon, and/or the allocation of gold certificates to, individual Reserve Banks. This would enable each Reserve Bank to increase its member bank reserve liabilities step by step with the Federal Reserve Bank of New York if credit conditions should demand a policy of ease. Alternatively, it would be necessary for the System

to make available to each Reserve Bank the securities required for open market sales, should credit conditions warrant a policy of greater stringency. Of course, the chief impediment in the way of bringing about such an institutional change is the difficulty of establishing regional capital markets of sufficient absorptive capacity, though progress along this line could doubtless be made if such a policy were adopted by the Federal Reserve System.

The power to change reserve requirements has obvious potentialities in the present context. The exercise of this authority has usually been relegated to circumstances calling for a "basic" or "structural" change in the banking community's reserve position. More frequent use has been avoided on the ground that the instrument is too blunt for the purposes at hand. On the other hand, a change in reserve requirements has an immediate and far-reaching impact upon the commercial banking system. Serious consideration should be given to the provision of a more important role for this regulatory device, if the elimination of lags in the transmission of the effects of monetary policy is to be achieved. With present authority extended to a greater range, lags according to bank class could be eliminated. Differences in the degree of credit tightness at a given time which result from the correspondent relationship could be erased by the same means. Moreover, this instrument of control could be given the desired degree of fineness by simply reducing the magnitude of the changes decreed.[8]

Of the traditional instruments of central bank control, rediscount rate policy remains. With minor exceptions, changes in bank rate are uniformly made in all Federal Reserve districts. And since rediscount rate changes are more effective as indications of policy shifts than as means of directly controlling the flow of credit, little could be expected from their more frequent use.

Lest the above paragraphs be miscontrued as condoning or favoring the formation, as well as the execution, of monetary policy on a regional basis, it should be stated that nothing of the sort is envisaged. Rather, in order to avoid embarrassing inconsistencies, which have occurred during the history of the Federal Reserve System, monetary policy should continue to be determined at the Federal Open Market Committee level. Furthermore, the greatest care

8. *Cf.* Charles R. Whittlesey, "Reserve Requirements and the Integration of Credit Policies," this *Journal*, LVIII (Aug. 1944), 559–60.

Those who fear the direct price or yield effects of open market operations should be the ones most easily persuaded of the merits of more extensive use of changes in reserve requirements.

would have to be exercised in order to avoid the use of regional controls as a sop to local interests. Otherwise, regional monetary policy might offset the local credit effects of interregional flows of trade and resources.

In conclusion, it is true that institutional changes, such as the development of a correspondent banking system, the establishment of the Federal Reserve System, the growth of financial intermediaries operating on a national scale, and the growth and dispersion of the federal debt, have succeeded in destroying many of the vestiges of regional compartmentalization in the capital market. However, in view of the empirical findings cited in this study, the suggested modifications of our present system of central banking controls would undoubtedly improve the speed and efficiency of monetary policy.

<div align="right">

IRA O. SCOTT, JR.
</div>

HARVARD UNIVERSITY

ACCELERATED DEPRECIATION:
SOME FURTHER THOUGHTS

By ROBERT EISNER*

Recent articles in this *Journal* and elsewhere have indicated that, under the assumption that the stream of gross investment is growing — an assumption which is not unreasonable, either for the representative American firm or for the economy as a whole — accelerated depreciation has the somewhat surprising effect of increasing permanently the annual flow of depreciation charges.[1] Accelerated depreciation may thus affect income in several distinct respects.[2]

First, it transports depreciation charges for given assets on the average to an earlier period of time. Hence, for firms (or economies) whose investment is not continuous, and where taxes are computed on income net of depreciation charges, it reduces near future *taxable*

1. Evsey D. Domar, "The Case for Accelerated Depreciation," this *Journal*, LXVII (Nov. 1953), 493–519, and "Depreciation, Replacement and Growth," *Economic Journal*, LXIII (Mar. 1953), 1–32; Robert Eisner, "Accelerated Amortization, Growth and Net Profits," this *Journal*, LXVI (Nov. 1952), 533–44, and "Depreciation Allowances, Replacement Requirements and Growth," *American Economic Review*, XLII (Dec. 1952), 820–31. See also Eric Schiff, "A Note on Depreciation, Replacement, and Growth," *Review of Economics and Statistics*, Feb. 1954, pp. 47–56, and George Terborgh, *Realistic Depreciation Policy* (Chicago: Machinery and Allied Products Institute, 1954).

2. Since this comment was first written, the issues to which it is directed have been sharpened by passage of new federal tax legislation. The Internal Revenue Code of 1954 (H.R. 8300) provides for substantial acceleration of depreciation. It allows specifically for tax purposes the declining balance method, with a rate as much as twice the straight line rate, and the even faster sum of the years-digits procedure. By this latter system, in which cost is allocated in proportion to the terms of an arithmetic progression, the first of which is equal to the assumed number of years of life of the property and the last of which is unity, almost three quarters of the cost of each asset may be charged to depreciation during the first half of its life. (For example, with a property lasting thirty years, depreciation charges in the first year would be thirty divided by the sum of the digits one to thirty, or 30/465, which equals approximately 6.5 per cent. The depreciation reserve would exceed half of the original cost of the property by the end of nine years. In the first fifteen years, a reserve for depreciation of 345/465, over 74 per cent of the original cost of the property, would accumulate.) The new tax law, which is applicable to capital expenditures on or after January 1, 1954, will, in accord with the general mathematical conclusions of both Domar and myself, involve a sharply increased volume of depreciation charges. A discussion focused directly on the nature and implications of the accelerated depreciation provisions of the 1954 tax law may be found in Robert Eisner, "Depreciation under the New Tax Law," *Harvard Business Review*, Jan.–Feb. 1955, pp. 66–74.

income and raises distant *taxable* income. Second, during the very substantial period of transition to accelerated depreciation (that is, while regular depreciation charges on older properties are added to the accelerated depreciation charges on new assets, a matter of thirty or more years for many firms), accelerated depreciation raises depreciation charges and reduces taxable income.[3] Third, in the case of firms (or economies) where the stream of gross investment is growing, it raises depreciation charges, reduces taxable income and hence reduces corporate income taxes *permanently* (for as long as the growth assumption obtains). To the extent that firm policy is influenced by the accountants' reports of "net taxable profits," there may be, in each of the respects above, temporary delays, permanent postponements, or permanent reductions in extra-firm disbursement of earnings — to dividend collectors, research consultants, advertising agencies and employees, as well as to the tax collector.

In my own work I have hitherto concentrated on the direct tax and distributional aspects of the relationships involving depreciation policies and the phenomenon of growth.[4] Professor Domar has very wisely extended the discussion beyond these direct effects to implications of the relationships for economic growth and development.[5] In his most recent published work he has suggested specifically that accelerated depreciation would "orient" tax policy towards the ends of "economic growth" and "the development of new firms."[6] This brief note will question the merit of that suggestion. In so doing it will explore some relatively neglected implications of depreciation policies for the theory of income and employment.[7] It will emphasize in particular that the tax considerations which Domar offers as significant in the decisions of the individual firm tell only a small part

3. With a given tax structure the transition to accelerated depreciation *delays* corporate income taxes. It should be noted, however, that while in the case of a firm whose capital expenditure programs are discrete, these delays in the payment of income taxes are temporary (though recurrent with each new wave of expenditures), in the case of a firm with more or less continuous expenditures the delay in tax payments is in effect permanent, or for as long a period as the firm exists (continues to replace exhausted properties).

4. In addition to articles cited above, see my "Conventional Depreciation Allowances Versus Replacement Cost," in *The Controller*, XXI (Nov. 1953), 513–14, 533, and my *Rejoinder* to a comment in the *American Economic Review*, XLIII (Sept. 1953), 614–21.

5. See his articles cited in n. 1, above.

6. This *Journal*, LXVII, 493.

7. Domar apparently attempts to rule out many of these implications by a questionable assumption that each firm could be induced to "keep a special set of books for income tax purposes" and thus "avoid confusion in wage, dividend and other policies" (p. 500). (See below, pp. 293–94.) .

of the story and hence lead to improper conclusions when we consider the problem at the level of the economy as a whole.

As indicated above, I have no question in regard to the fundamental arithmetic and algebraic relations which depict the level of depreciation charges as a function of economic growth and the speed of depreciation. Domar's case for accelerated depreciation founders, however, on the improper deduction that since (1) accelerated amortization raises depreciation charges (and lowers taxes) if the stream of gross investment is growing, therefore (2) accelerated amortization will encourage "economic [aggregate?] growth."

While admitting that "ideally, we should want to know the effects of accelerated depreciation on investment decisions" and inferring that for this purpose "existing investment theory . . . is . . . inadequate," Domar bases his entire "case for accelerated depreciation" on an unsubstantiated assumption, or as he puts it, "the hope that a high ratio ["between accelerated and normal depreciation allowances"] will be conducive to investment and to development of new firms" (p. 494). Let us see what objective basis we can furnish for this "hope."

The *partial* derivative of investment with respect to acceleration of depreciation may be considered positive for the following reasons:

(1) In the period of transition to accelerated depreciation, taxes are reduced. If the firm "dies," or stops making capital expenditures, the tax reductions are merely deferrals. More generally, the tax savings are retained as long as the firm continues to make capital expenditures, that is, indefinitely. These tax savings may be used for capital expenditures (as well as for other expenditures or merely for improving the firm's liquidity or cash position). In regard to new firms (or new investment) accelerated depreciation delays disbursement of funds to tax collectors (and others). It thus makes available to the firm interest free *loans* which *may*, as above, be used for purchase of capital equipment.

(2) In the important case of growing firms where, as Domar declares, "A growing firm means a firm with a growing stream of investment," accelerated depreciation, by actually raising the volume of depreciation charges (and not merely postponing a given volume), makes *additional income* available for capital expenditures. Given the inevitable imperfections in the capital market, the opportunity cost of so utilizing such income is likely to be less than the cost of securing external funds for capital expenditure purposes.

(3) Whether a firm is growing, stationary or declining, accelerated depreciation, by increasing expected near future incomes after

taxes on each unit of investment at the expense of incomes expected after the accelerated period is completed, raises the marginal efficiency of investment schedule. (This effect is enhanced to the extent that the firm discounts future returns for reasons of individual time preference, which include discounting for risk and uncertainty.)

(4) In growing firms, acceleration of depreciation raises the marginal efficiency of investment schedules by raising the entire expected level of profits after taxes and thus improving the general climate of expectations (to the extent continued growth is expected and the relation between growth and depreciation rates is understood).

The explicit statement above of the probable effects of accelerated depreciation on the investment of the firm may immediately suggest to the reader the rather substantial logical gap between these probable effects and the increased rate of growth desired by Domar. For Domar states, "Throughout this paper, growth is expressed exclusively in terms of investment in fixed capital. A growing firm means a firm with a growing stream of investment" (p. 494). Yet his argument, if valid at all, would indicate only a rise in the level of investment and not an increase in its rate of growth. In other words, Domar can claim at most that accelerated amortization will raise the first derivative of capital stock with respect to time but not its second derivative.[8]

However, how confident can we be that acceleration of depreciation will raise even the level of *aggregate* investment? A conclusion that the *partial* derivative of investment with regard to the acceleration of depreciation is positive could apply on the level of the individual firm, or perhaps even in regard to a geographical area or industry in need of stimulus; but it is entirely improper, in looking at the economy as a whole, to ignore the indirect effects on investment stemming from the influence of accelerated depreciation on other elements in the determination of the level of national income. It must, of course, be recognized that if accelerated depreciation for tax purposes is to be introduced while all other elements of fiscal policy — tax rates as well as government expenditures — are kept unchanged, income, under reasonable assumptions, is virtually certain to rise. But this would follow from *any* decrease in taxes and cor-

8. The reader should realize that this is a basic issue and no mere mathematical quibble. Domar himself has pioneered in showing that long-run full employment is a problem of attaining not merely a high level of investment but of attaining a high rate of growth in the level of investment. (See, for example, his "Expansion and Employment," *American Economic Review*, XXXVIII, March 1947, 34–55; also my own "Underemployment Equilibrium Rates of Growth," *American Economic Review*, XLII, Mar. 1952, 43–58.)

responding budget deficit. What are the implications of accelerated depreciation that would serve to differentiate it from any of a number of other fiscal measures for raising the aggregate level of income and, in varying degree, the components of that income?

To the extent that depreciation charges are higher, the claims on output of the various factors of production must be lower. The immediate short-run effect as viewed by the accountant may be only a reduction in net "profits." But the nature of the bargaining process involving the various claimants to "national income" is such that when "profits" are lowered both dividends to stockholders and employee incomes are likely to be reduced. And if, as is frequently argued, depreciation charges influence price policy, higher depreciation charges reduce real income of the factors of production by as much as higher corporate receipts resulting from higher prices are not passed on to the factors of production. A strong case may certainly be built to the effect that higher depreciation charges (such as would result from accelerated depreciation in a growing economy) reduce the consumption component of aggregate demand.[9]

But it is hardly reasonable to ignore the effect upon the marginal efficiency of investment schedule of such a lowering of the consumption function.[1] In fact, it appears likely that arguments (3) and (4),

---

9. *The General Theory of Employment Interest and Money* may well be reread in this regard. Keynes writes: "We must not underestimate the importance of the fact . . . that, whereas employment is a function of the expected consumption and the expected investment, consumption is, *cet. par.*, a function of *net* income, *i.e.*, of *net* investment (net income being equal to consumption *plus* net investment). In other words, the larger the financial provision which it is thought necessary to make before reckoning net income, the less favourable to consumption, and therefore to employment, will a given level of investment prove to be" (pp. 98–99). Keynes observes further that depreciation allowances and sinking funds "are apt to withdraw spending power from the consumer long before the demand for expenditure on replacements (which such provisions are anticipating) comes into play; *i.e.* they diminish the current effective demand and only increase it in the year in which the replacement is actually made. If the effect of this is aggravated by 'financial prudence,' *i.e.* by its being thought advisable to 'write off' the initial cost *more* rapidly than the equipment actually wears out, the cumulative result may be very serious indeed" (p. 100).

1. If higher depreciation charges do in fact entail higher prices, there may be a direct effect on the marginal efficiency of investment schedule. It might be argued that current supply prices of capital goods would be increased by accelerated depreciation more than would expected product prices, affected as the latter might be by uncertainty (perhaps even including uncertainty as to the permanence of accelerated depreciation and high supply prices of capital goods). This would involve a lowering of the marginal efficiency of investment schedule even without introduction of the consumption effects we discuss. However, I am inclined personally to attach such a high order of uncertainty as to the various hypothetical contingencies involved in this argument as to be unwilling to do more than call it to the reader's attention, without endorsement.

above, indicating a positive sign for the partial derivative of investment with respect to acceleration of depreciation, or the total derivative for the individual firm, will lose much of their strength when applied to the economy as a whole. If accelerated depreciation applies throughout the economy, each firm may find the gain in lessened taxes in part offset by the loss in gross receipts.[2] One may be struck by the similarity of the argument that increased depreciation charges will increase investment to the now well-recognized fallacy of composition that if *all* individuals (and firms) try to save more, more saving (investment) will actually take place.[3]

A complete consideration of the effects of the acceleration of depreciation should involve the issue of substitution of capital equipment for labor or other factors of production. It should be recalled that expenditures for labor services, which are used up entirely, presumably, in the current year's output, are entirely deducted from the current year's receipts in calculating taxable income. (Indeed, long-run and institutional advertising expenses which result in increased sale of output, increased prices of output, and/or lower costs of marketing output in future years are also deducted entirely from receipts of the year in which they are incurred.) Acceleration of depreciation would offer firms an incentive to spend more on capital equipment and, to the extent that various expenditures are substitutes for one another (as alternative means of making possible increased sales and/or lower total costs — including market-

---

2. The reader should beware of a likely fallacious objection to my arguments. In this objection one would declare that accelerated amortization would raise depreciation charges only if new investment were in fact undertaken. With no investment there would be no depreciation charges. Hence the problem of lower consumption or declines in other noninvestment expenditures would only arise to the extent that the purpose of the accelerated depreciation — more investment — were being realized.

But this argument is fallacious because the higher depreciation charges would apply to *all* investment, including investment which would have been undertaken without additional tax incentives. Domar touches upon this problem of added depreciation charges which may have little effect on current investment decisions in his recommendation of high initial allowances as opposed to a mere acceleration of "straight-line" depreciation (this *Journal*, LXVII, 503, 508). Higher depreciation charges which initially induce investment will involve a greater drag on the economy in subsequent periods. But higher depreciation charges on the investment which would have been undertaken anyway represent a net drag on the economy in the periods of investment as well.

3. The extent to which Domar seems to have confounded individual and aggregative investment theory is further evidenced by his citation of George Terborgh's *Dynamic Equipment Policy* (*ibid.*, 494, n. 6) as an exception to his statement about the inadequacy of existing investment theory. For Terborgh's excellent work is devoted almost exclusively to the theory of investment for the individual firm!

ing costs — of sales), less on current services. In addition to cancelling out at least some portion of the added contribution to the income stream stemming from increased investment as a result of accelerated amortization, this substitution factor might reasonably be expected to affect the distribution of income in a fashion unfavorable to consumption by decreasing the price of labor services (which would, of course, be the major substitute for capital goods).[4] In this expectation we would, of course, be assuming a higher marginal propensity to consume out of labor income than out of income from capital. Yet, while thus tending to lower the aggregate demand function, substitution of investment expenditures for, let us say, advertising expenditures or somewhat padded expense accounts for salaried employees would tend to increase productive capacity. However, this would mean an increase in capacity with a change in type of expenditure but no increase in total expenditure. Hence we would have an increase in *excess* capacity and in unemployed or underemployed resources. One may then conjecture reasonably that with such a development of excess capacity it would prove increasingly difficult to maintain a high and growing level of investment.

We may well have here a fallacy of composition over time; that is, acceleration of depreciation, even if it raises investment initially will run into the problem that the long-run demand for investment, that is, the amount of investment which will not prove redundant, depends upon the long-run level of aggregate demand and the long-run rate of growth in the level of aggregate demand.[5] If these are not affected favorably by accelerated depreciation we cannot, except for relatively small movements from consumption goods production to investment goods production, expect a favorable effect on investment.

An important factor which must not be overlooked is the rela-

4. The added intervention in the economy occasioned by accelerated depreciation is, of course, not neutral in respect to allocation of resources and is indeed probably less neutral than fiscal policies which involve, for example, increasing the exemptions on taxable personal income (regardless, of course, of the source of the income). This lack of neutrality should not be overlooked but I am not personally of the philosophy that all added "non-neutral" governmental intervention is bad and I am not prepared to evaluate this particular intervention from the standpoint of resource allocation. By reducing firms' dependence upon external funds for financing investment, however, accelerated amortization, regardless of the particular direction of governmental intervention, reduces the influence of the market in the allocation of capital and hence impairs the long-run workings of competitive forces.

5. Cf. Keynes, "New capital-investment can only take place in excess of current capital-disinvestment if *future* expenditure on consumption is expected to increase. Each time we secure to-day's equilibrium by increased investment we are aggravating the difficulty of securing equilibrium tomorrow" (*The General Theory*, p. 105).

tionship between replacement requirements and depreciation allowances. For if depreciation allowances exceed replacement requirements even the investment of all of these allowances as they accrue (and no net investment in the accounting sense) will involve an expansion in capacity. But the more rapid is growth the greater the excess of depreciation charges over replacement requirements and the more acceleration of depreciation will increase this excess.[6] It should be noted also that in new firms, since relatively few assets have had a chance to get worn out, replacement requirements also tend to be exceptionally low in regard to depreciation charges. Now, it should be recognized that for a host of reasons, institutional as well as theoretical (e.g., a rationally justifiable tendency to expect current demand to continue rather than to expect it to expand), replacement expenditures are likely to receive readier approval than capital expenditures which could only be justified by other considerations.[7] Yet it is in precisely those firms — growing firms and new firms — where the problem of securing a high level of investment demand reduces itself to that of utilization of large proportions of gross saving in nonreplacement expenditures, that acceleration of depreciation, by increasing gross saving, would aggravate the problem.[8]

This note certainly does not pretend to be an exhaustive discus-

6. Cf. my articles in the *American Economic Review* (Dec. 1952), and this *Journal* (Nov. 1952) and Domar's in the *Economic Journal* (Mar. 1953), all cited above. However, Domar ignores this relationship in his article in this *Journal*.

7. Cf. Joel Dean, *Capital Budgeting* (New York: Columbia University, 1951) and Terborgh's *Dynamic Equipment Policy;* also my "Interview and Other Survey Techniques and the Study of Investment," to be published in Vol. XIX of *Studies in Income and Wealth*, and the works that I cite there.

8. Keynes sees in this problem an important factor in the great American depression following the crash of 1929: "In the United States, for example, by 1929 the rapid capital expansion of the previous five years had led cumulatively to the setting up of sinking funds and depreciation allowances, in respect of plant which did not need replacement, on so huge a scale that an enormous volume of entirely new investment was required merely to absorb these financial provisions; and it became almost hopeless to find still more new investment on a sufficient scale to provide for such new saving as a wealthy community in full employment would be disposed to set aside. This factor alone was probably sufficient to cause a slump. And furthermore, since 'financial prudence' of this kind continued to be exercised through the slump by those great corporations which were still in a position to afford it, it offered a serious obstacle to early recovery" (*The General Theory*, p. 100). How different is the import of this quotation from that of Domar's argument that increased depreciation allowances, as brought on by accelerated depreciation, "will be conducive to investment" and "economic growth"! For, of course, the Keynesian system has no room for the *assumption* that increased saving by particular individuals or categories of individuals (or firms) will be conducive to investment. It is surprising that Domar, whose own work has offered such a stimulating extension of the Keynesian system, should have failed at least to point up his reversal of doctrinal field in this regard.

sion of issues raised by the proposal of accelerated depreciation. Any full consideration should underscore the point that at best, the cyclical influences of this device promise to be highly perverse. In a period of boom and particularly of price inflation, according to Domar's arguments, acceleration of depreciation would stimulate firms to increase investment expenditures. To do this would intensify inflationary pressure, with full employment assumed. It would also intensify inflationary pressure even where, as we suggest, firms transfer expenditures from labor services, which would provide output available for the currently inflated demand, to investment goods or labor services for the production of investment goods, which would not bear fruit until some future period (possibly of economic downswing or depression). In a downswing or at the end of a depression, accelerated depreciation might actually give *lower* depreciation charges than normal depreciation methods which would involve charges on the earlier assets purchased during the period of boom. However, even if the accelerated depreciation did not result in lower depreciation charges (because the dip was not long or sharp enough) the general fall in profits (if not of tax rates) in a depression would vitiate any tax saving effects from accelerated depreciation.[9]

It is also important to have a clear notion that, on the one hand, depreciation allowances are, aside from tax considerations, merely a paper entry; they do not in themselves actually give the firm any funds. Yet, on the other hand, it would be naive to deny that they influence firm policies with respect to dividend payments and wage negotiations, if not product pricing as well, which do effect the funds at the disposal of the firm. Domar's position in this regard is somewhat ambiguous. At one point he declares: "In order to avoid confusion in wage, dividend and other policies, it is suggested that in addition to its regular books where normal depreciation is recorded, each firm should keep a special set of books for income tax purposes."[1] This would indicate that firms would somehow be expected to ignore the legally accepted depreciation charges in formulating company policies — a pious hope, the chances for the realization of which must seem slim indeed.[2] Yet, in his Table II and surrounding dis-

9. Domar has recognized the cyclical problem, at least in part. Cf. this *Journal*, LXVII, 509–10.

1. *Ibid.*, p. 500.

2. One may observe the considerable extent to which companies have utilized special sets of books without Internal Revenue Service recognition where such books, by listing higher depreciation charges, make "profits" appear lower and hence put management in an easier position for retaining earnings and resisting demands for higher wages. Higher depreciation charges mean that management is laying a greater "claim" on the gross income of the firm. It would seem that

cussion Domar treats "depreciation charges and retained profits" as essentially additive.[3] Thus he assumes various rates of "retained profits" net of depreciation and calculates "the ratio of depreciation charges and [plus] retained profits to gross investment" for various rates of growth of gross investment and depreciation periods of twenty and thirty years. But for any given level of gross investment the firm whose rate of growth of gross investment has been *higher* would have *lower* depreciation charges than the firm with the slower rate of growth (of gross investment). Unless we are assuming that the firm with higher depreciation charges lowered wages or raised prices sufficiently to preserve the same rate of profit *after* depreciation deductions as the firm with the smaller depreciation deductions one would be quite unwarranted in various implications that might carelessly be read into Table II in its current form.[4]

Finally, I should like to note that calling into serious question the role of accelerated amortization in encouraging economic growth, the development of new firms, and a high level of investment, suggests that its effects on income distribution should not be casually dismissed.[5] There are a variety of fiscal measures which might much more reasonably be expected to accomplish the objectives indicated. The insistent pressure in certain business and government circles for widespread application of accelerated depreciation may therefore understandably lead to the thought that the dismissed distributional considerations are really paramount. For it is politically much

once the tax authorities recognized higher depreciation charges for tax purposes, management would have a much stronger incentive than at present to make these higher charges effective instruments of a very real claim on an increased share of the gross revenue of the firm.

3. *Ibid.*, pp. 497–99. These items could be additive only if firms did in fact relate actions to depreciation charges and thus evidence "confusion in wage, dividend and other policies." For unless higher depreciation charges led companies to decrease wages or dividends or to raise prices they would involve merely corresponding reductions in retained profits.

4. Thus, for example, Domar's Table II indicates that a firm with a 1 per cent rate of retained profits and a 6 per cent rate of growth of gross investment would be able to finance 65 per cent of its gross investment internally if the depreciation period were 20 years but only 55 per cent if the depreciation period were 30 years. But shortening the period of depreciation for the assets of any given firm would certainly not accomplish such an increase in the proportion of investment which could be financed internally if the firm kept "a special set of books" so as to inhibit any possible effects on firm policies. The amount of funds available internally would not be increased unless the firm took action to maintain its accounting rate of profit in spite of the higher resultant depreciation charges. (I am not charging Domar himself with falling explicitly into this error. His Table II and surrounding discussion, however, leave the reader without warning quite at the edge of the trap.)

5. Cf. Domar, this *Journal*, p. 509 (including n. 7).

more expedient to affect the incidence of taxation in a regressive fashion by this "subtle" method of redefining taxable income, about which the average voter cannot be expected to be well-informed, than to alter existing tax rates in favor of corporate income. It would be unfortunate if economists were not clear as to the relevant economic issues, no matter what public policies their political preferences may lead them to favor.

A Comment on Mr. May's Article.

In his "Concepts of Business Income and Their Implementation," this *Journal*, February, 1954, Mr. George O. May has referred to my work in a manner which requires some factual correction. Mr. May declared (p. 15) that (1) I was among those who "have argued that existing [1952] depreciation provisions are excessive," and (2) I "have proposed readoption and extension of the methods . . . by which provision for property exhaustion was deferred until replacement actually took place or became imminent." The record of my writings in this subject indicates however that:

(1) Far from arguing that depreciation charges have been excessive in the sense that May's unqualified usage implies, I had explained: "The term 'excessive' is employed here merely to indicate the arithmetic relation to replacement requirements. No implications are intended as to desirable policy from the point of view of financial reserve, investment decisions or other such matters."[6] As noted below, I do not myself argue that replacement requirements are an adequate criterion for depreciation charges. I insist only that those who do so argue note that, when proper allowance is made for the factor of growth, one may well have to conclude: "To the extent that replacement requirements may offer a criterion for the size of depreciation allowances . . . perhaps depreciation allowances are too high and net profits, as well as net income and net investment, are understated by conventional accounting practices!"[7]

(2) I nowhere have made any such "proposal" as May attributes to me. Rather, I declared specifically "I should classify myself as agnostic on the issue of whether the cost of replacing 'worn out equipment' (or 'the assets used up in production') should be a criterion for the adequacy of depreciation charges."[8] In fact, subsequently I devoted myself to an exposure of some of the fallacies of the "replacement cost" appeal.[9]

6. *American Economic Review*, Dec. 1952, p. 824, n. 4.
7. *Ibid.*, Dec. 1952, p. 831.
8. *Ibid.*, Sept. 1953, p. 616.
9. *The Controller*, XXI (Nov. 1953).

One small point which requires particular correction is May's puzzling ascription to me of the "reflection that under depreciation accounting all costs for renewal or replacement during the life of the unit are charged against maintenance."[1] The reader may note in my original article "the reflection that depreciation charges do not include charges for maintenance costs."[2] I have never stated anywhere that all costs for renewal or replacement are charged against maintenance.

NORTHWESTERN UNIVERSITY   ROBERT EISNER.

A COMMENT

By GEORGE O. MAY

On Professor Domar's Article.

In a note to my article, which appeared in the February issue of this *Journal*, the editor drew attention to the article, "The Case for Accelerated Depreciation" by Professor E. D. Domar, which had appeared in the November, 1953 issue.[1] He has offered me the opportunity to make brief comments on that article from a practical standpoint.

1. I suggest that the postponement of liability to income taxation for an indefinite period by allowing charges which clearly relate to a future year to be taken as deductions in an earlier year, is an undesirable form of subsidy. The announcement that the British Chancellor has last year virtually abandoned the system of initial allowances described in Domar's article, and has substituted a frank subsidy, indicates that he holds the same view. His action apparently met with general approval.

2. The British system of initial allowances differs in essentials from the American provisions for *accelerated amortization* under Section 124A of the federal tax law to which Domar refers.

The British allowance was made only from year to year and varied in maximum rate from 40 per cent to zero. Domar recognizes this fact and states that in principle such a variation is desirable. The American provision was established for a period defined in the law and fixed as to rate.

The British allowances were designed to encourage investment that was deemed economically desirable. The American provisions applied only to expenditures which were certified as being necessary

1. This *Journal*, Feb. 1954, 16 n.
2. *American Economic Review*, Dec. 1952, p. 823.
1. LXVIII, 10, n. 8.

or desirable for defense purposes, that were in many cases economically inopportune.

3. Domar states that if an anticipation of depreciation is regarded as a loan, "this loan need never be repaid unless the investment stream declines, and further, . . . [in] a growing stream — the most realistic American case — will result in a permanent reduction in the effective tax rate." (*sic.*) The suggestion that the loan need never be repaid (more accurately perhaps will be replaceable by a new loan) is unwarranted so long as there is no assurance that the system of allowances will be maintained at least at the same level.

4. Though Domar speaks of "the most realistic case," there are a number of types of cases, all of which are equally realistic. Domar's equation is best fitted to the type of industry which is well-established and can look forward to continuous growth. It is not well fitted to the enterprise which is still in its early stages. It is not well adapted to the enterprise which involves a large initial outlay and only relatively unimportant additions thereto, such as the Empire State Building. It is not well adapted to the enterprise which has to build up its business after it has created its facilities. It is admittedly not well adapted to the enterprise that has seen its best days, like the majority of our railroads. And it is not well adapted to our regulated public utilities, whose income taxes are treated as a cost to be recouped from consumers.

5. Turning from realism to relative importance, the National Income Supplement of 1951[2] estimated depreciation allowances for 1947 at approximately $12 billion. Of this total, slightly under $3 billion went to unincorporated owners of real estate; more than $1.4 billion went to utilities; allowances to manufacturing corporations were $2.35 billion, or less than 20 per cent of the whole. And in the five years from 1948 to 1953, investment in construction was nearly as great as investment in producers' durable equipment.

6. Domar recognizes that it would be preferable to deal separately with classes of property having widely different life expectancies, but he does not stress this point adequately. Authorities agree that at least three classes should be recognized: long, medium and short-lived.

Depreciation charges are allowed in respect of property having a useful life of five years or less at the one extreme and seventy-five years or more at the other. Since property having long life carries a relatively low rate, and since the average rate is, on the basis of

2. Department of Commerce, National Income Supplement, 1951, p. 138 and underlying data.

Domar's assumption of an average life of thirty years, only 3⅓ per cent, it is clear that the great bulk of depreciable property is in the medium or long-life classifications.

7. The word "depreciation" has a well-established meaning consistently assigned to it in our income tax law since 1918; viz., "a reasonable allowance for the exhaustion, wear and tear, and obsolescence of property." The standard method of measuring this exhaustion from year to year is, no doubt, a process of amortization. But it does not follow, as Domar seems to assume it does, that any amortization system is a depreciation system.

Accelerated depreciation is a term used to describe depreciation which is deemed to be more than normally rapid because of abnormalities in the extent or conditions of use. What Domar calls "accelerated depreciation" is at best an anticipation of depreciation expected to be sustained in the future.

The allowances under Section 124A of our law (to which Domar refers as the American "system") are rightly described in the statute as *accelerated amortization* and are explicitly offered as an alternative to depreciation. The British initial allowance is obviously not even an amortization allowance.

On Professor Eisner's Articles.

Much of the foregoing has a bearing also on Professor Eisner's articles and on his complaints immediately above (p. 285). His first criticism of my reference to him, as I see it, is that I put too favorable an interpretation on what he said. The second is that I attributed to him a view which he had expressed and which was quite correct. If, in the circumstances, I owe him an expression of regret, I gladly offer it.

1. Upon his own interpretation of his article, "Perhaps depreciation allowances are too high . . ." ("Summary and Conclusions," *American Economic Review*, Dec. 1952, p. 831), and its corollaries are left without the slightest support in argument, authority or evidence, and his article is left barren of any useful conclusion. Further, the statistics cited in paragraph 5 of my Comment on Domar show that the continuous growth of individual firms is not as decisively significant or as completely characteristic of our economy as he has assumed in that article (at p. 826) and in his article in the November 1952 issue of this *Journal* (p. 538).

2. Under his definition of depreciation, it is legitimate to assume, as he did, that keeping the asset that is being depreciated in good operating condition until its retirement (which, of course, includes making any necessary replacements or renewals of parts)

is effected "by appropriate repair and maintenance expenditures" (p. 823). Under a different system, replacements and renewals of important parts would be charged against depreciation provisions. (See my article, p. 10.)

General Comment

The provisions of the Revenue Act of 1954, which permit the use of the diminishing-balance method of depreciation, reflect a view that existing depreciation allowances may be too low. It may safely be said that (quite apart from changes in price levels) straight-line rates that were appropriate thirty years ago would be inadequate today.

The real criticism of the straight-line method is that it gives the allowance *earlier* (in some cases much earlier) than they are needed; the annuity method has been proposed, and in some cases adopted, to compensate for this defect. On the other hand, straight-line rates have been criticized as inappropriate because the loss of usefulness is usually greater (in some cases much greater) in the earlier than in the later years; on this ground, the diminishing-balance method has been favored by some.

During the last thirty years, the lowering of rates of interest on industrial borrowings has lessened the force of the first argument, while the great increase in obsolescence that has accompanied the increase in investment has added to the force of the second.

Southport, Connecticut   George O. May.

A REJOINDER
By Evsey D. Domar

It is not easy to unravel Professor Eisner's thoughts, but in addition to comments on traps, on doctrinal purity, and on the beguiling of the electorate (pp. 288, 291–95), he seems to make the following main points:

Accelerated depreciation will:

(1) possibly increase the level of investment but not its rate of growth (p. 288);

(2) enlarge retained earnings at the expense of dividends, wages and other expenditures (pp. 289–91);

(3) redistribute income in favor of capitalists and against labor and other factors (pp. 286, 289, 290–91, 295);

(4) in view of (2) and (3) lower the consumption function and thus produce a deflationary effect which may reduce even the level of investment (p. 289);

(5) make capital redundant relative to the long-run rate of growth of demand and create unused capacity (p. 291);

(6) substitute capital for labor and other factors with subsequent unemployment and reinforcement of the undesirable distributive effects listed in (3) (pp. 290–91);

(7) possibly reduce the marginal efficiency of capital (pp. 290–91), and presumably the volume of investment, by raising prices, particularly of capital goods;

(8) intensify the business cycle (pp. 292–93).

It begins to look like the list of grievances in the Declaration of Independence. It did not occur to me that a seemingly innocent tax proposal such as accelerated depreciation could be so potent a force — in the wrong direction. But since there is nothing sacred about so-called normal depreciation, and since most, if not all, of these arguments are reversible, a system of *decelerated* depreciation should be just as potent, and, of course, in the right direction. Limitation of space must have prevented Eisner from expounding the virtues of this novel and marvelous fiscal weapon.

Before we take up Eisner's list, let us examine the heart of the issue. Accelerated depreciation affects the firm in two ways. First, with given income tax rates, it postpones and frequently reduces the firm's income tax payments. Second, it usually results in a lower net profit, both before and after taxes, *on the firm's books*. The first effect is real, in the sense that it actually improves the firm's position and prospects irrespective of its accounting procedures.[1] The second effect is essentially an accounting paradox and can be corrected by formal entries on the books or merely by a mental note on the part of management. To evaluate Eisner's arguments, it is important to distinguish between these two effects.

My next point is that any *workable* tax proposal must be a more or less blunt instrument for the simple reason that workability (at least in democratic countries) implies a rather broad application without excessive administrative discretion. Many appealing tax proposals must therefore be abandoned in favor of theoretically less satisfying alternatives. This bluntness almost invariably gives rise to undesirable by-products and even abuses. An increase in personal income tax exemption as a means of redistributing income and stimulating consumer expenditures during a deflation will, I trust, be acceptable to Eisner; yet a $100 increase will put some $20 into the pocket of a relatively poor person (and none at all into that of the

---

1. Under some circumstances, accelerated depreciation can be unwelcome to a taxpayer; for instance, if he expects tax rates to rise, if he has a fluctuating income without adequate carry-over and carry-back provisions, and sometimes if he is subject to a progressive income tax. Hence, it should not be made compulsory.

very poor), as compared with the $70 or even $90 accruing to the very rich, of which a good part may not be spent at all.

Accelerated depreciation is no exception to this rule; it was suggested in my paper as a method of relieving new and growing firms from a part of their income tax burden simply because I did not then, and do not now, know of any reasonably practical *direct* method of achieving this objective. If and when a better method appears, I shall be very happy to trade. Eisner suffers from no such ignorance; he knows that "there are a variety of fiscal measures which might more reasonably be expected to accomplish the objectives indicated" (p. 294), and it is regrettable that he chose not to enlighten us with a single example.

Let us now turn to point (1) on Eisner's list. It is entirely possible that accelerated depreciation will increase neither the rate of growth of investment, nor, for that matter, its level. Certainly, no guarantees can be given. All we can do is to study the nature of a tax proposal, observe it in operation (if possible) and compare it with alternatives. As far as the level and growth of investment are concerned, accelerated depreciation should be superior at least to a corresponding reduction in the corporate income tax rate because the tax relief afforded by this treatment of depreciation goes only to those firms who are investing (in fixed capital); because the relief is temporary and peters out unless subsequent investments are made; and finally, because the greatest advantage is enjoyed by those firms whose investment is growing. In fact, accelerated depreciation can redistribute business income from nongrowing to growing firms. As with any tax proposal, all these attributes may still fail to produce the desired results, and accelerated depreciation was presented in my paper as "just an instrument, one of many, and . . . certainly not . . . a panacea." (p. 510). Since Eisner has presented no alternative proposals, further discussion along these lines is beside the point.[2]

Let us proceed to points (2), (3) and (4). Eisner is perfectly correct in stating that the tax relief and any consequent increase in saving created by accelerated depreciation should be compared not with the total volume of investment, but only with the increment which can be ascribed to this change in the tax laws. But is this not true of any measure? For a tax proposal to merit consideration, it is not necessary to show that it is superior to all others, under all possible circumstances, and no such claim for accelerated deprecia-

2. I can, however, but marvel at the analytical virtuosity which allows him to argue that accelerated depreciation is not likely to increase the amount of investment in the aggregate, and yet to list the replacement of labor by the additional capital constructed in response to this measure among the latter's defects — all this in one paper.

tion was made. Nor would anyone argue that if a given reduction in tax revenues is contemplated, the whole amount must necessarily take this form. All that is really required is to show that accelerated depreciation, as a form of tax relief, is for certain purposes (e.g., for encouragement of investment and development of new firms) superior to *at least* one other reasonable alternative, such as a straight reduction in the corporate income tax rates. This, I hope, my paper did show, though here again I would not argue that accelerated depreciation wins even this limited contest on all possible occasions. It may, for instance, lead to great abuses arising from the sale of secondhand assets, such as old buildings, because of the low rates of our capital gains tax.

Eisner seems to be aware that accelerated depreciation, being a form of tax relief, cannot be deflationary by itself unless counteracted by other tax increases or corresponding reductions in government expenditures (pp. 288–89). Here again there is little use arguing about the final effect unless the nature of these offsetting measures, if any are taken, is specified. But the awareness of this fact does not diminish Eisner's insistence that accelerated depreciation must be deflationary.

Much of his opposition to this measure stems from its accounting attributes, namely, from the fact that larger depreciation charges and hence smaller book profits will lower dividends, wages and other expenditures, and even raise prices. Studies by Brown on depreciation methods and by Butters on inventory profits have shown that the effects on business decisions of accounting calculations as such vary widely from firm to firm and that no simple generalization can be made.[3] It is difficult to believe, however, that having won the long advocated privilege of rapid amortization, business would then forget about the lower tax payments, greater liquidity and higher *real* profits, and behave as though it were worse off — all because of an accounting paradox.[4] The strict conformance of depreciation charges to allowances under the tax laws has not been universal, to put it mildly, and discussions among accountants show that they

3. E. Cary Brown, *Effects of Taxation: Depreciation Adjustments for Price Changes* (Boston, 1952); J. Keith Butters, *Effects of Taxation: Inventory Accounting and Policies* (Boston, 1949).

4. Lower tax payments, greater liquidity and higher real profits will be enjoyed, at least for a while, by every corporation investing in fixed capital if accelerated depreciation is not accompanied by a rise in the corporate income tax rates. Otherwise, only new and rapidly growing firms will be better off, while others will pay higher taxes. All this abstracts from secondary effects of these tax changes.

are quite aware of the problem created by the five-year amortization period allowed under the certificates of necessity issued during the war.[5] Hence my suggestion that accelerated depreciation should be recorded for tax purposes only, may not be just a "pious hope." In addition, labor unions are not unable to interpret accounting statements; and it is highly questionable whether our price mechanism is so irrational that a mere change in recording depreciation will raise prices in spite of higher real profits. Yet Eisner's argument has some substance. Much will depend on whether depreciation is sufficiently accelerated to make it worth while to take account of the difference between it and the normal charges; there is no use denying that the measure included in the recent revision of our tax laws is much too mild to pass this test. Nor is the new treatment of depreciation radical enough to have appreciable effects on investment decisions. It was a compromise, but of that unfortunate variety which shares in the defects of both extremes but not in their virtues.

Points (5) and (6) regarding the redundancy of capital and its substitution for labor apply to any investment, whether or not motivated by accelerated depreciation, and hence are outside the scope of this discussion. Eisner must know very well that the solution to his problem lies in growth; he may be reminded that investment not only increases capacity and replaces labor when the construction of new capital is finished, but that it also employs labor during the construction period and generates income via the multiplier process. If he really believes in what he says, and if in addition he objects to new firms on the amazing ground that their fixed capital need not be replaced for some time, as if new capital bought by established firms were any different (pp. 291–92), then he has the strongest and the best argument — from his point of view, of course — against accelerated depreciation which makes all his other arguments entirely superfluous. It is a waste of time to discuss the merits of a proposal whose objective is regarded as undesirable in the first place.

Point (7), which deals with a possible rise of the prices of capital goods and hence with a fall in the marginal efficiency of investment, requires no comment because Eisner, having made it, disclaimed any responsibility for it. May I suggest, however, that if uncertainty is as important as he thinks (pp. 289–90), which may very well be true, then

5. See for instance the discussion of the "Restoration of Fixed Asset Values to the Balance Sheet," *The Accounting Review*, XXII (April, 1947), 194–210; also William A. Paton, "Measuring Profits under Inflation Conditions: A Serious Problem for Accountants," *The Journal of Accountancy*, LXXXIX (Jan. 1950), 16–27.

accelerated depreciation acquires an additional virtue because, by postponing tax payments until a good part of the investment has been recovered, it surely reduces uncertainty. His treatment of point (8) regarding the cyclical aspects of accelerated depreciation is rather puzzling, because he does no more than repeat my own argument, minus the remedy suggested. We need not deprive him of the fun of breaking through open doors, except to mention that if the accounting effects of accelerated depreciation are indeed as strong as he asserts, then the resulting understatement of profits during prosperity (when new investments are made and written off) and their overstatement during depression should have a welcome stabilizing effect.

Underlying this disagreement about accelerated depreciation there is, I believe, another and a deeper one, shared by many others. For good or for bad, our economy is endowed with a positive net propensity to save, technological progress, and a growing population — a combination which makes growth in output, and very probably in investment as well, an indispensable condition for the preservation of full employment. If for one reason or another an adequate volume of investment is not forthcoming, the economy has a choice between reducing the propensity to save, thus achieving full employment at a lower rate of growth, or realizing the full growth potential by stimulating investment. One need not take an extreme position on every occasion, but the difference in emphasis will remain. Under more normal circumstances, the choice between these alternatives would be an ordinary value judgment, one, which in spite of its apparent simplicity, is not easy to make because a growing society has many other attributes besides an increasing output. But the present international situation has, I believe, eliminated this choice and left us no alternative but to grow and grow rapidly, though strictly speaking, the preference for self-preservation also involves a value judgment.

Turning now to Mr. May's comment, I regret that he regards accelerated depreciation as "an undesirable form of subsidy." Since no reasons for this opinion are given, no further comment on my part is called for. His point (3) is very well taken; I should have made it explicit in my paper. Finally, I like his suggestion that the measure under discussion should be called accelerated amortization rather than depreciation.

<div align="right">Evsey D. Domar.</div>

The Johns Hopkins University

PROFESSOR SAMUELSON ON OPERATIONALISM
IN ECONOMIC THEORY

By Donald F. Gordon

Professor Samuelson's *Foundations of Economic Analysis* has, as one of its main objectives, to demonstrate the existence of operational propositions in economic theory, and to show how they can be derived. By an operational proposition, Samuelson means a proposition which implies an "operation" of some kind whereby, in principle, at least, it can be tested; even though, for practical or financial reasons, it may be impossible to carry out such an operation in fact. It is the purpose of this comment to show that his methods are in all cases not sufficient to establish operational propositions, in some cases are not necessary, and in some are either meaningless or positively misleading.[1]

Professor Samuelson begins by pointing out that the setting up of sufficient conditions or functions to determine a number of unknown economic variables is not, by itself, producing an operational theory. That a demand curve and a supply curve determine price and sales, or that the equation of marginal revenue and marginal cost determines output, are not, for example, operational propositions. Economists have proceeded to develop what may be called "multipliers" in a generalized sense; that is, they have derived algebraic expressions for the rates of change of dependent variables associated with changes in independent variables, or associated with shifts in some of the functions involved. These multipliers are in terms of the slopes of the original equations, and in many cases the original hypotheses setting up these functions specify something about the signs of their slopes. Even here, however, it may not be possible to determine so much as the sign of the multiplier, for it may include too many of these slopes for a determinate answer.[2]

1. Despite the fact that the original version of the *Foundations* carried the subtitle "The Operational Significance of Economic Theory," Samuelson in the present work is concerned with many other topics besides operationalism. While the present author believes the methodology is largely in error, the fascinating insights and brilliant techniques displayed by Professor Samuelson have put all economic theorists in his debt, and many of his treatments will be the starting points in the discussion of their respective problems for years to come.

2. For example, a relatively simple Keynesian type of multiplier is $dY/dM = C_i + F_i/[L_y(I_i + C_i) + L_i(1 - C_y - I_y)]$, where Y, I, C, i, and M are national income, investment, consumption, the interest rate and the amount of money,

But Samuelson believes that the signs of these multipliers may be derived from one of two fundamental hypotheses, which, he believes, are the "foundations of economic analysis" in the sense that they are the foundations of any operational propositions which can be derived therefrom. First we have the hypothesis of maximization or minimization, where the original equilibrium functions are necessary conditions for the maximization or minimization. In these cases, the secondary conditions for the maximum or minimum determine certain properties of the slopes of the equilibrium functions, and these in many cases will determine, in turn, the signs of the multipliers. These can then be tested by observation.[3] The second hypothesis is that of dynamic stability. We relax the static equations to allow for nonequilibrium positions, but hypothesize that in such cases, the variables approach the equilibrium positions over time. For this hypothesis to hold, certain properties of the slopes of the functions must hold, and these properties may determine, in turn, the signs of the multipliers. Again, the signs of the multipliers will be testable by observation. This latter method Samuelson has termed the Correspondence Principle, since it demonstrates a relationship between comparative statics and dynamics.[4] More than the qualitative signs of the multipliers can hardly be expected from pure theory alone, without further quantitative data.

Our first criticism of this methodology is that the two foundations are, at the very least, not sufficient. If one of the functions in the theory shifts, then the observed changes in dependent variables may not coincide with changes predicted from the signs of the multipliers, as derived from the two fundamental hypotheses. If this happens, there are two possibilities. The shift may be predictable in the sense

---

respectively; and L_i, L_y, I_i, I_y, C_i and C_y are marginal propensities to hold money, invest, and consume, with respect to the interest rate and national income respectively. Even knowing the signs of these slopes and something of their orders of magnitude does not determine the sign of the whole multiplier.

3. We may illustrate the method with a simple example adapted from Samuelson where the result is obvious. (Cf., *Foundations*, pp. 14–16.) If a firm has a total revenue function, $R(x)$, and a total cost function $C(x) + tx$, where x is output and t is the amount of a tax per unit, then the necessary condition for profit maximization is that $R'(x) - C'(x) - t = 0$. The multiplier effect of an increase in the tax per unit is $dx/dt = 1/[R''(x) - C''(x)]$, which is negative since the secondary condition for profit maximization is that the denominator is negative.

4. To take, again, the simplest possible case, where the result is well known, take a Keynesian system consisting of a consumption function $C(Y)$ and investment I, with the ordinary multiplier, $dY/dI = 1/1 - C'$. $|C'| < 1$ is a condition for a dynamic system involving a single period lag between C and Y to be stable. If, therefore, we hypothesize stability, the multiplier is positive.

that it may be related to some change in another observable variable. If this is true, either the generalization must be restricted to situations where this observable variable does not change; or the observable variable must be incorporated as another independent variable in the function that shifted, and a new theory derived. On the other hand, if the function is "truly" unstable in the sense that its shifts must be ascribed to a change in "tastes," etc., then the conclusion will simply be false. Thus an additional foundation of economic analysis will be the hypothesis that the functions used are not unstable in this sense. (In what follows "unstable" will mean unpredictably unstable, and "stable" will not exclude predictable shifts.)[5] Thus there does not appear to be any way, even with Samuelson's technique, to avoid the tedious, detailed, empirical work necessary to establish the additional foundation.

Our second criticism grows out of the first and is concerned with Samuelson's attempt to derive the multiplier effects of "shifts" in functions. If this is restricted to predictable shifts the procedure is legitimate, although the multiplier effects are really those of the independent variables from which the shifts can be predicted. On the other hand, if these are unpredictable shifts, then the procedure has no meaning; for these shifts make impossible the derivation of any operational propositions from the fundamental hypotheses — even those multipliers relating to shifts in observable independent variables. Consequently there is no point to Samuelson's analysis of the effects of a shift in a function due to a change in "tastes," or a shift which is presumably an autonomous change in behavior patterns since it is left unexplained.[6]

In the third place, the two hypotheses may be, from a methodological point of view, not even necessary. If a theory is composed of stable functions, then arbitrary or controlled shifts in the independent variables — which are necessary to test any operational hypothesis of these types — will trace out functional relationships between the independent and dependent variables. From these, not only qualitative, but quantitative results can be obtained, without at any time considering explicitly either of the fundamental hypotheses. In this connection it would be well to notice a qualitative difference in the

5. Needless to say, "stable" here has quite a different meaning than it has when an equilibrium is described as "dynamically unstable." Practically, of course, "stable" cannot mean precisely stable.

6. T. C. Schelling, who uses Samuelson's Correspondence Principle in his *National Income Behavior* (New York, 1951), likewise analyzes the effects of unexplained shifts in the parameters of functions. Cf., p. 104.

two fundamental hypotheses. In the case of maximization or minimization, the hypothesis itself is the first step in deriving the original equations determining the equilibrium values of the variables. Even if the secondary conditions for the maximum or minimum are not explicitly considered, it could not be said the hypothesis is unnecessary. On the other hand, in the case of the hypothesis of dynamic stability, it is quite unnecessary to consider it at any time, since it is not crucial to the derivation of the original equilibrium equations.

A fourth criticism also concerns only the fundamental hypothesis of dynamic stability, and concerns its validity, rather than the derivation of propositions from it. Recent theories of the business cycle, as well as various forms of the cobweb theorem, suggest that actual economic variables may possess *no* stable equilibrium values over the observable range, yet the values observed may all be points on stable functions. In these cases the hypothesizing and observation of stable functions will give correct quantitative slopes for these functions, while the hypothesis of dynamic stability will actually lead to incorrect values even qualitatively. Defending this second hypothesis, Samuelson writes,

> The plausibility of such a stability hypothesis is suggested by the consideration that positions of unstable equilibrium, even if they exist, are transient, nonpersistent states, and hence on the crudest probability calculation would be observed less frequently than stable states.[7]

But this implies that what we actually observe are either points of dynamically stable or nonstable equilibrium, which, if true, would make the stable situation by far the more probable. What, in fact, we may be observing, and would be, in the cases cited above, are neither.

A final criticism which we believe can legitimately be made of Samuelson's discussion of the Correspondence Principle relates not to the validity of the method, but to the minds of economists. As noted above, he correctly points out that the formulation of sufficient functional relationships to determine a number of unknowns accomplishes nothing by itself. In the simplest case of a demand and supply curve, if all he could say was that there were "efficacious laws" determining price and output, "The economist would be truly vulnerable to the gibe that he is only a parrot taught to say 'supply and demand.' "[8] Yet he continues,

> "In the above illustration let us consider 'tastes' as a changing parameter influencing only the demand curve. Will an increase in demand raise or lower

7. *Foundations*, p. 5.
8. *Foundations*, pp. 257–58.

price? Clearly the statement that before and after the assumed change price is determined by the intersection of supply and demand gives us no answer to the problem. Nothing can be said concerning the movement of the intersection point of *any* two plane curves as one of them shifts. And yet most economists would argue that in a wide variety of circumstances this question can be given a definite answer — namely, that price will increase.

"How is this conclusion derived? For few commodities have we detailed quantitative empirical information concerning the exact forms of the supply and demand curves even in the neighborhood of the equilibrium point. Not only would large amounts of time and money be necessary to get such information, but in many cases it is practically impossible to derive useful empirical information concerning what would happen if virtual changes in price confronted the demanders or the suppliers."

If the second point listed above is correct, operational propositions derivable from this type of shift are impossible; but we are here chiefly interested in another matter. Samuelson proceeds to dynamize the model, and, from the necessary conditions for dynamic stability, deduces that price will rise. Very likely many economists have been surprised to find that what "little success" they have had in this question has been due to the Correspondence Principle: nevertheless, Samuelson is undoubtedly correct in feeling that most economists would give a definitive answer of some kind in this situation. That conditions for dynamic stability are not the basis for such an answer can be seen by recalling that an answer to this type of question was given many decades before demand and supply functions, let alone dynamic stability conditions, were used. Moreover, a similar answer is given today by businessmen who may be quite unaware of either device. To go back no further than Adam Smith, starting from a position of equality in price and cost, he writes:

"When the quantity of any commodity which is brought to market falls short of the effectual demand, . . . the market price will rise more or less . . . according as either the greatness of the deficiency, or the wealth and wanton luxury of the competitors, happen to animate more or less the eagerness of the competition."[9]

The purpose in this quotation is not to indulge the human propensity for uncovering an "ancient" source of a modern subtlety. It is to point out that this is a quite different proposition than one about the effects of a shift in a demand *curve*. And it is about this one that we feel certainty. What Smith is saying is that a discrepancy between a *quantity* demanded and a quantity supplied will produce a rise in price, an observation which must have been made since free markets have operated. It is surely because of this latter proposition that we

9. *Wealth of Nations*, Cannan Edition, p. 50.

feel certainty about the effects of a change in tastes (quantity demanded), rather than because the conditions for dynamic stability indicate a price rise. Smith's proposition is not inconsistent with demand and supply curves, but by itself is a qualitative dynamic proposition requiring no function at all.

<div align="right">Donald F. Gordon.</div>

University of Washington
Seattle

<div align="center">COMMENT</div>

<div align="center">*By* Paul A. Samuelson</div>

Professor Donald F. Gordon has listed five criticisms of the methodology of my *Foundations of Economic Analysis*. To the extent that I understand his arguments, it is primarily his fourth criticism concerned with the Correspondence Principle relating dynamic stability and comparative statics that seems to me to be in need of amplification and qualification. Here are my reactions to the specific points Gordon makes. I hope that these hasty interpretations will be regarded as tentative.

Criticism 1. The hypothesis that a competitive firm with an unchanging cost curve will never reduce its quantity supplied when market price rises is according to my use of terms a meaningful proposition: i.e., it is conceptually an empirically refutable proposition. I continue to aver that the hypothesis of profit maximization is sufficient to deduce this hypothesis.

Gordon's criticism seems to me to be concerned not so much with the sufficiency of my reasoning but rather with the quite different problem of how we go about deciding that a conceptually meaningful proposition is or is not a fruitful hypothesis and whether as a result of any particular observations we are to decide that it has or has not been refuted.[1] This is an important question: indeed much of the whole theory of statistical inference is concerned with little else, and in the last decades due to the work of Haavelmo, Koopmans, Frisch, Elmer Working, and other writers on the problem of "identification" a small start on a satisfactory theory has been made. I do not think that I discussed this issue anywhere in *Foundations*, and in retrospect I feel little urge to have done so. To test, refute, or "verify" a meaningful proposition is a tough empirical job and no

---

1. E.g., a fall in quantity supplied by a maximizing competitor may have been due to a simultaneous factor price increase.

amount of flossy deduction can obviate this. Realizing this should not be disillusioning; and it certainly should not tempt one to belittle meaningful propositions in favor of empty ones.

Criticism 2. This does overlap with the first point. Again Gordon seems to me to be saying: "In real life, observed changes are often the result of simultaneous changes in numerous parameters; but just which parameters have changed or how great are the relative quantitative magnitudes of the changes is a most difficult question to answer. Therefore, it is wrong (or misleading? or useless?) to try to set up meaningful hypotheses about the *ceteris paribus* effects of changes in each datum taken separately." I do not think the second sentence follows from the truth of the first.

There is one special point in this second criticism that may require special comment. I do indeed think there is point in analyzing changes in "tastes." How will thriftiness affect capital formation? How will a shift from beer to tea affect markets? How will knowledge on the part of the consumer that liver is good for anemia affect the relative prices of kidneys and liver? These are perfectly legitimate questions to ask, whether or not we have an "explanation" for the change in tastes. Indeed, in a fundamental methodological sense, it is precisely the changes in the parameters or data of a system that are, within the framework of *that* system, unexplainable. I do not see that I, or Schelling, or Hicks (some of whose *Value and Capital* theorems refer to defined changes in tastes) have any reason to avoid trying to answer such questions. I only wish that we had better success in doing so.

Criticism 3. To me this says that there may be ways of forming meaningful hypotheses other than by postulating maximization or dynamic stability. Of course, there are. No one has ever doubted it.

Observe market behavior over time; make statistical scatter diagrams; and if the result suggests to you the hypothesis that the marginal propensity to consume is exactly .925, or that the elasticity of demand for rye is $-.70$, or that the propensity to save schedule is concave from above — then well and good, for these are all meaningful, refutable hypotheses. (The last four sentences of Gordon's criticism deal with a different issue and not one that requires any comment here.)

Criticism 4. The point made here seems to me to be a very fruitful one. The Correspondence Principle is a vague line of deductive reasoning by which we can under certain circumstances deduce, from the hypothesis that a stipulated system is dynamically stable, various

implied hypotheses about its comparative statical properties appropriate to a permanent shift in a parameter to which corresponds a new stationary equilibrium position. There is much that is unsatisfactory about the exposition of this heuristic principle and about our knowledge of its deductive properties. But the point that Gordon is here concerned with is a different one. In *Foundations* I did not content myself with deriving these formal properties: in addition, I stepped forward as a man of the world and casual empiricist and stated my opinion that the hypothesis of dynamical stability was a "realistic" one to make. I am no longer so sure of this.

True, there is something vaguely persuasive about the doctrine of "the nonpersistence of unstable states" that Gordon quotes from my book: indeed, as my reference to L. J. Henderson will show, this is not a new thought originated by Keynes to refute criticisms made in this *Journal* concerning the properties of the marginal propensity to consume; instead it can be traced back to Hippocrates, to Maxwell, to Gibbs, to Darwin, and to a host of philosophers; and when the Smyth Report told us that a certain sized arrangement of Uranium 235 would explode, most of us amateurs used this heuristic reasoning to predict that such concentrations of Uranium 235 were not likely to be found in nature by geologists.

None the less, you never get something for nothing and never empirical hypotheses from empty deductive definitions. At best your observation can tell you only that the real world (or some subset of it) is not exploding; your theoretical model or system will always be an idealized representation of the real world with many variables ignored; it may be precisely the ignored variables that keep the real world stable, and it takes a significant act of inductive inference to rule this out and permit the Correspondence Principle to deduce properties of the idealized model.

Here is an example. Some critics of Keynes alleged that he did believe, and needed to believe, that the consumption function is convex. Closer reading of his book shows this not to be so. Keynes replied arguing that all that he required was a marginal propensity to consume of less than one. And even if this were not granted, he felt that this would simply throw on the doubter the need to explain why the capitalistic system is not hopelessly unstable.[2]

2. My discussion of the Correspondence Principle should have made clearer the possibility that the world is stable but that our model — in this case a simple income model — is too simple to portray the world correctly. Thus, the marginal propensity to consume might exceed unity, but tightness of the money supply and movements of interest rates might be what prevents explosion. This raises questions of empirical judgment.

Well, maybe the system *is* unstable. That is one possibility, and as Gordon is cogently pointing out, many of the cobweb cycles and auto-relaxation trade cycle theories of such moderns as Kaldor, Goodwin, Hicks, and others are squarely based on the notion of a system that is locally unstable at its stationary levels so that it oscillates — but because of such nonlinear elements as full-employment ceilings, capacity limitations, impossibility of disinvesting faster than at certain limiting rates, the system oscillates with a preferred finite amplitude.

A priori reasoning will not settle this empirical question. During the late 1930's I felt, and so perhaps did writers like Professor Lloyd Metzler, that the observed behavior of the macroaggregates of the American economy was most compatible with the hypothesis of dynamic stability: in those days it seemed hard rather than easy for the system to generate self-aggravating cumulative movements of explosive type. This view may have been wrong, and we may have been too confident in expressing it.[3] Or it may have been a fruitful hypothesis for that period but not one for the present decade. Or it may still be fruitful. These are important empirical questions that cannot be answered by dividing dichotomously the world's possibilities into categories of stable and unstable and inferring that our observed world by its not having exploded away is necessarily in the stable category. So I concur in Gordon's misgiving.

Criticism 5. Here the question is whether older economists used the so-called Correspondence Principle in arriving at comparative statical results. And did they know they were using it? To the latter question the answer is surely, No. They could not have known the name, obviously; and they were not so self-conscious as to analyze the logical structure of many of their intuitions. But were they, so to speak, talking prose all their lives without knowing it?

I have never seriously examined the writings of Adam Smith and other economists to see just how much of the logic involved in the Correspondence Principle can by sympathetic reasoning be imputed

3. Readers of the modern cycle literature may receive the misleading impression that economists started with linear models and then arrived at nonlinear ones. The historical patterns are the reverse: most of the literary models that antedate the work of Frisch, Tinbergen, Metzler, and myself were very definitely nonlinear. One of the things we thought worth showing was the fact that no separate theory of the turning points or of the four phases of the cycle was needed: the same linear model could be shown to be theoretically capable of producing all phases of the cycle. This does not deny that there may be some truth to Haberler's gibe that with their nonlinear models the econometricians are almost catching up to the literary economists.

to them. I therefore welcome Gordon's quoted passage from Smith and will give it brief mathematical formulation:

$$(1) \qquad dp/dt = k(D - S)$$

where k is a positive number that may be approximated by a constant and where D is the quantity demanded (as shown by the demand schedule) and S is the quantity supplied (as shown by the supply schedule). To parrot Gordon's words, "Very likely Adam Smith would be surprised to find that he talked in terms of differential equations as well as prose." But, of course, how we choose to describe the contents of his thought will not alter its content or be relevant to his terminology.

Granting equation (1), do we conclude that it supplants or contradicts the Correspondence Principle? Not at all. From (1) alone, Smith and Gordon cannot conclude anything about *comparative statics*. They can make the following *comparative dynamic statement:* "If price has been stationary, and if after time t_0 the demand curve shifts rightward (and remains there), then in the following time intervals we can expect price to rise from its initial level." That is all.

If now we want to make a comparative static statement, we must somehow add hypotheses. Thus, we might stipulate the empirical hypothesis that such a shift in demand will always be followed by an approach to a new stationary equilibrium. This is an hypothesis of dynamic stability: and with this hypothesis, we can infer from (1) that the comparative statical price change must be positive. But in so doing we are using — guess what? Of course, the Correspondence Principle.[4]

In conclusion, may I repeat my lack of confidence that I have correctly apprehended Professor Gordon's friendly criticisms.

Paul A. Samuelson.

Massachusetts Institute of Technology

---

4. An even better case is provided by Gordon's earlier example of the simple multiplier model: $y_t = c \, y_{t-1} + I$, where all variables are measured in deviations from some equilibrium levels. Our primitive instinct is dynamic: a higher marginal propensity to consume c or I will raise money income. But if c is already greater than one, the comparative statical theorem tells us that a permanent rise in autonomous investment or consumption spending is compatible only with a *lower* level of income! Again, it is the Correspondence Principle that is involved, only this time on the pathological side.

EXPECTED PRICE AND THE COBWEB THEOREM

By Arnold P. Collery

A fundamental assumption of the cobweb theorem is that pro-
ducers expect that the price in the coming period will be the same
as the present price. Although this may be an adequate description
of price expectations in some competitive markets at certain periods
of time, it need not be true. Many other situations are equally con-
ceivable. This note will explore the implications of seven alternative
assumptions.

(1) Producers may assume that if the price has changed between
the last two periods, it will change in the *same* direction by (a) a
greater amount, (b) the same amount, or (c) a smaller amount.

(2) Producers may assume that if the price has changed between
the last two periods, it will change in the *opposite* direction by (a) a
greater amount, (b) the same amount, or (c) a smaller amount.

(3) Producers may assume that the price will be an average of
past prices.

Let:

$$P \ = \ \text{price}$$
$$P^e = \ \text{expected price}$$
$$t \ \ = \ \text{time period}$$
$$D \ = \ \text{amount demanded}$$
$$S \ = \ \text{amount supplied}$$

Making use of these symbols, all of the above cases can be
generalized into the following equation.

(1) $\qquad P^e(t) = P(t-1) + K\,[P(t-1) - P(t-2)]\,.$

The reader can easily verify that $K > 1$ corresponds to Case (1a)
above, $K = 1$ to Case (1b), $1 > K > 0$ to Case (1c), $K < -1$ to
Case (2a), $K = -1$ to Case (2b), $0 > K > -1$ to Case (2c), and
$K = -.5$ to Case (3). If we set $K = 0$ we obtain the original
assumption of the theory, i.e., $P^e(t) = P(t-1)$.

Assuming linear demand and supply equations, our demand
equation can be written as:

(2) $\qquad D(t) = a - b\,P(t)\,;$

our supply equation as:

(3) $S(t) = c + h P^e(t)$.

Substituting equation (1) into equation (3) we obtain:

(4) $S(t) = c + (h + hK) P(t - 1) - hK P(t - 2)$.
Setting $S(t) = D(t)$ and solving for $P(t)$ we obtain:

(5) $P(t) = (a - c)/b + (-h - hK)/b\, P(t - 1) + hK/b\, P(t - 2)$.

Solving this difference equation and examining its roots, we derive the following diagram. On the horizontal axis positive values of K are plotted to the right of the origin and negative values to the left. On the vertical axis is plotted the absolute value of the ratio of the slope of the supply curve (h) to the slope of the demand curve (b), i.e., h/b .
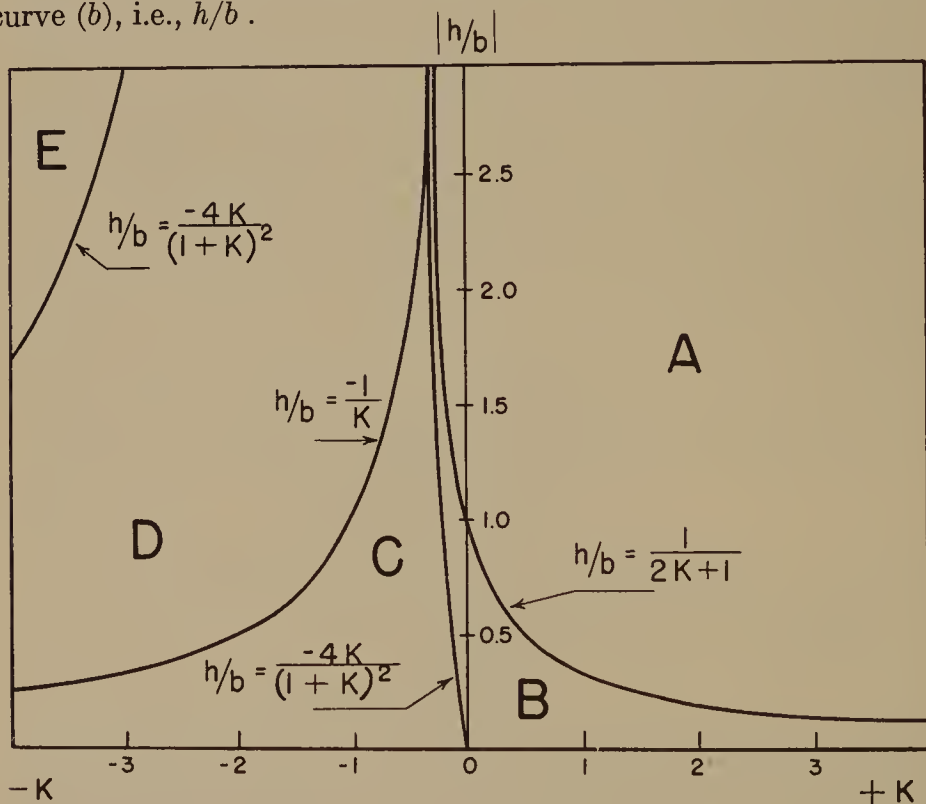
In Area A are shown all combinations of $|\,h/b\,|$ and K which result in two-period divergent cycles.[1]

1. A two-period cycle is a fluctuation in price such that the price is in one period above its equilibrium value, the next period below it, and then above it again. Thus a four-period cycle would be one where price is above its equilibrium value for two consecutive periods, then below it for two periods, and then above it again.

In Area B are shown all combinations of $|h/b|$ and K which result in two-period convergent cycles.

In Area C are shown all combinations of $|h/b|$ and K which result in four-period convergent cycles.

In Area D are shown all combinations of $|h/b|$ and K which result in four-period divergent cycles.

In Area E are shown all combinations of $|h/b|$ and K which result in price exploding in either direction.

The conclusions of the original theorem can be found on the vertical axis, i.e., where $K = 0$. Thus when $|h/b| < 1$ we are in Area B and price approaches equilibrium. If this slope ratio > 1, we are in Area A and price fluctuates further and further from equilibrium.

The diagram can be used to show what would happen if, given the slopes of the curves, the suppliers change the basis on which they anticipate future price. If, e.g., $|h/b| = 1$ and $K = 0$, we have undampened, two-period cycles; this point lies on the boundary line between Areas A and B in the diagram. If the producers learn from experience and shift, for example, to Case (2c) and stick there, the point in the diagram which represents this combination of K and $|h/b|$ will be either in Area B or C, since we have assumed that $|h/b| = 1$ and $0 > K > -1$. Thus price would eventually approach equilibrium.

<div style="text-align: right">ARNOLD P. COLLERY.</div>

AMHERST COLLEGE

RECENT PUBLICATIONS

ADAMS, GEORGE P., JR. Competitive Economic Systems. New York: Thomas Y. Crowell Company. 1955. pp. 516. $6.00.

ALLAIS, M. Les Fondements Comptables de la Macro-Economique. Paris: Presses Universitaires de France. 1954. pp. 91.

AMERICAN PHILOSOPHICAL SOCIETY. Ideology and Reality in the Soviet System. Philadelphia, Pennsylvania: American Philosophical Society. 1955. pp. 38. $1.00. [Proceedings, Volume 99, No. 1.]

BAUER, P. T. West African Trade: A Study of Competition, Oligopoly and Monopoly in a Changing Economy. New York: Cambridge University Press. 1954. 450 pp. $10.00.

BELCHER, ALGERNON S. A Study of the Factors Determining the Exercise of Jurisdiction by the National Labor Relations Board in Three Selected Industries. Washington, D. C.: Catholic University of America Press. 1954. pp. 62. [Studies in Economics, Abstract Series Volume No. 10.]

BOUVIER, EMILE. Le Droit du Travail au Quebec. Montreal, Canada: Les Editions Bellarmin. 1954. pp. 31.

BROOKINGS INSTITUTION. Economics and Public Policy: Brookings Lectures, 1954. Washington, D. C.: The Brookings Institution. 1955. pp. 157. $2.00.

CALCATERRA, ERCOLE. Il Rapporto fra i Livelli Nazionali dei Prezzi. Milan, Italy: Dott. A. Giuffre. 1953. pp. 200.

CARUS-WILSON, E. M. Essays in Economic History. New York: St. Martin's Press, Inc. 1954. pp. 438. $10.00.

CHARNES, A. AND COOPER, W. W. The Stepping Stone Method of Explaining Linear Programming Calculations in Transportation Problems. Pittsburgh, Pennsylvania: Carnegie Institute of Technology. 1954. pp. 69.

CLEGG, H. A. General Union: A Study of the National Union of General and Municipal Workers. Oxford, England: Basil Blackwell. 1954. pp. 358.

CLELAND, SHERRILL. The Influence of Plant Size on Industrial Relations. Princeton, New Jersey: Industrial Relations Section, Department of Economics and Sociology, Princeton University. 1955. pp. 65. $2.00.

COLM, GERHARD. Essays in Public Finance and Fiscal Policy. New York: Oxford University Press. 1955. pp. 375.

CONFERENCE ON RESEARCH IN INCOME AND WEALTH. Short-Term Economic Forecasting. Princeton, New Jersey: Princeton University Press. 1955. pp. 506. $7.50. [Studies in Income and Wealth, Volume 17.]

CONGRESS OF INDUSTRIAL ORGANIZATIONS. The Case Against "Right to Work" Laws. Washington, D. C.: Congress of Industrial Organizations. 1955. pp. 171. $2.00.

CONGRESS OF INDUSTRIAL ORGANIZATIONS, DEPARTMENT OF EDUCATION AND RESEARCH. A Handbook on State and Local Taxes. Washington, D. C.: Congress of Industrial Organizations. 1954. pp. 142. $2.00.

CROCKETT, JOSEPH P. The Federal Tax System of the United States: A Survey of Law and Administration. New York: Columbia University Press. 1955. pp. 288. $4.75.

DALTON, HUGH. Principles of Public Finance. London: Routledge and Kegan Paul, Ltd. 1954. pp. 255. [Fourth Edition.]

DE NEUMAN, A. M. Industrial Development in Indonesia. Cambridge, England: Students' Bookshops, Ltd. 1955. pp. 40.

DICKEN, SAMUEL NEWTON. Economic Geography. Boston, Massachusetts: D. C. Heath and Company. 1955. pp. 569. $6.50.

EDELMAN, MURRAY. National Economic Planning by Collective Bargaining: The Formation of Austrian Wage, Price, and Tax Policy after World War II. Urbana, Illinois: Institute of Industrial Relations, University of Illinois. 1954. pp. 78.

EDITORS OF FORTUNE. The Changing American Market. Garden City, New York: Hanover House. 1955. pp. 304. $4.50.

FEROLDI, FRANCO. Formazione del Capitale e Politica Creditizia nello Sviluppo Economico. Milan, Italy: Dott. Antonino Giuffre. 1954. pp. 118.

FISHER, IRVING. The Theory of Interest. New York: Kelley and Millman, Inc. 1954. pp. 566. $8.50.

FISHER, ROBERT MOORE. The Metropolis in Modern Life. Garden City, New York: Doubleday and Company, Inc. 1955. pp. 401. $6.00.

FLANDERS, ALLAN AND CLEGG, H. A. The System of Industrial Relations in Great Britain: Its History, Law and Institutions. Oxford, England: Basil Blackwell. 1954. pp. 380.

FLIPPO, EDWIN B. Profit Sharing in American Business: A Study of Methods Used to Maintain and Sustain Profit-Sharing Plans. Columbus, Ohio: Bureau of Business Research, College of Commerce and Administration, Ohio State University. 1954. pp. 183.

FOWLER, CHARLES B., ET AL. Economic Handbook: A Visual Survey. New York: Thomas Y. Crowell Company. 1955. pp. 246.

GALBRAITH, JOHN KENNETH. Economics and the Art of Controversy. New Brunswick, New Jersey: Rutgers University Press. 1954. pp. 111. $2.50.

GALENSON, WALTER. Labor Productivity in Soviet and American Industry. New York: Columbia University Press. 1955. pp. 273. $5.50.

GENDARME, RENE. La Region du Nord. Paris, France: Librairie Armand Colin. 1954. pp. 250.

GOLDSMITH, RAYMOND W. A Study of Saving in the United States. Princeton, New Jersey: Princeton University Press. 1955. pp. 1769. $30.00. [Two Volumes.]

HECKSCHER, ELI F. An Economic History of Sweden. Cambridge, Massachusetts: Harvard University Press. 1954. pp. 308. $5.00. [Translated by Göran Ohlin.] [Harvard Economic Studies, Volume XCV.]

HENDERSON, HUBERT DOUGLAS. The Inter-War Years and Other Papers. New York: Oxford University Press. 1955. pp. 445. $6.75.

HUNOLD, ALBERT C. The Industrial Development of Switzerland. Cairo, Egypt: National Bank of Egypt. 1954. pp. 45.

INDUSTRIAL RELATIONS SECTION, PRINCETON UNIVERSITY. Selected References. Princeton, New Jersey: Industrial Relations Section, Princeton University. 1954. pp. 4. '20 cents. [No. 61.]

INTERNATIONAL COMMISSION OF JURISTS. For the Rule of Law. The Hague, Netherlands: International Commission of Jurists. 1955. pp. 39. [Bulletin No. 1.]

INTERNATIONAL ECONOMIC ASSOCIATION. International Economic Papers, No. 4. London and New York: The Macmillan Company. 1954. pp. 229. $3.50.

IRWIN, HAROLD S. Evolution of Futures Trading. Madison, Wisconsin: Mimir Publishers, Inc. 1954. pp. 95. $3.00.

ISAAC, JULIUS. British Post-war Migration. New York: Cambridge University Press. 1954. pp. 310. [National Institute of Economic and Social Research, Occasional Papers XVII.]

JEANNENEY, JEAN-MARCEL. Les Commerces de Detail en Europe Occidentale. Paris: Librairie Armand Colin. 1954. pp. 69. [Cahiers de la Fondation Nationale des Sciences Politiques, 61.]

JOHNSON, JOHN L. Income in Kentucky. Lexington, Kentucky: University of Kentucky Press. 1955. pp. 310. $4.75.

JOSHI, RAO BAHADUR N. S. AND DHEKNEY, B. R. Irrigation and Agriculture in the First Five Year Plan: An Appraisal. Poona, India: Deccan Book Stall. 1954. pp. 110. [Volume I.]

KERR, CLARK. The Trade Union Movement and the Redistribution of Power in Postwar Germany. Berkeley, California: University of California, Institute of Industrial Relations. 1955. pp. 29. 20 cents. [Reprint No. 62.]

LENT, GEORGE E. The Ownership of Tax-Exempt Securities, 1913–1953. New York: National Bureau of Economic Research, Inc. 1955. pp. 140. $1.50. [Occasional Paper 47.]

LETICHE, JOHN M. The System or Theory of the Trade of the World. Baltimore, Maryland: John Hopkins Press. 1954. pp. 24. $1.50.

MAVERICK, LEWIS A. Productivity: A Critique of Current Usage. Carbondale, Illinois: Lewis A. Maverick. 1955. pp. 30. 50 cents.

MENDERSHAUSEN, HORST. Two Postwar Recoveries of the German Economy. Amsterdam, The Netherlands: North-Holland Publishing Company. 1955. pp. 130.

MIERNYK, WILLIAM H. Inter-Industry Labor Mobility. Boston, Massachusetts: Bureau of Business and Economic Research, Northeastern University. 1955. pp. 158. $2.75.

MOLEY, RAYMOND. What Price Federal Reclamation? New York: American Enterprise Association, Inc. 1955. pp. 72. $1.00. [No. 455 in the series "National Economic Problems."]

NATIONAL MANPOWER COUNCIL. A Policy for Skilled Manpower. New York: Columbia University Press. 1954. pp. 299. $4.50.

NEUFELD, MAURICE F. Day In, Day Out with Local 3, IBEW. Ithaca, New York: New York State School of Industrial and Labor Relations, Cornell University. 1955. pp. 69. [Bulletin No. 28.]

NEVIN, EDWARD. The Problem of the National Debt. Cardiff, Wales: University of Wales Press Board. 1954. pp. 115.

PAULSEN, ANDREAS. Neue Wirtschaftslehre. Berlin, Germany: Verlag Franz Vahlen. 1954. pp. 387.

ROTHSCHILD, K. W. The Theory of Wages. Oxford, England: Basil Blackwell. 1954. pp. 178.

SOLO, ROBERT A. (EDITOR). Economics and the Public Interest. New Brunswick, New Jersey: Rutgers University Press. 1955. pp. 318. $5.75.

SWARUP, RAM. Communism and Peasantry. Calcutta, India: Prachi Prakashan. 1954. pp. 194. $5.00.

TAFT, PHILIP. The Structure and Government of Labor Unions. Cambridge, Massachusetts: Harvard University Press. 1954. pp. 312. $6.00.

UNITED NATIONS. Foreign Capital in Latin America. New York: Columbia University Press. 1955. pp. 164. $1.75. [U. N. Sales No. 1954.II.G.4.]

U. S. CONGRESS, JOINT COMMITTEE ON THE ECONOMIC REPORT. January 1955 Economic Report of the President. Washington, D. C.: U. S. Government Printing Office. 1955. pp. 1267.

WINDMULLER, JOHN P. American Labor and the International Labor Movement, 1940 to 1953. Ithaca, New York: Institute of International Industrial and Labor Relations, Cornell University. 1954. pp. 243. $3.00. [Cornell International Industrial and Labor Relations Reports, No. 2.]

WRONSKI, HENRI. Le Role Economique et Social de la Monnaie dans les Democraties Populaires. Paris, France: Librairie Marcel Riviere et Cie. 1954. pp. 181.

THE

QUARTERLY JOURNAL OF ECONOMICS

FOUNDED 1886

The Quarterly Journal of Economics

Published by Harvard University

Books, periodicals, and manuscripts to be addressed, THE QUARTERLY JOURNAL OF ECONOMICS, M-12, Littauer Center, Cambridge, Mass.

Business letters to be addressed, HARVARD UNIVERSITY PRESS, 44 Francis Avenue, Cambridge, Mass.

Subscriptions $5.00 a year. Three years, $14.00. Single copies (except those published in two parts) $1.50, postpaid. A special subscription rate of $3.00 is offered to students. Such special subscriptions may be for one, two, three, or four years. Those desiring to take advantage of this reduced rate should include with their remittance a statement from an officer of their college or university that they are *at that time* "regularly enrolled students who have not yet received the Doctor's degree."

CONTENTS FOR MAY, 1955

RECENT OPEN MARKET COMMITTEE POLICY
AND TECHNIQUE

By Deane Carson[*]

I. Introduction

Economists are in general agreement that open market operations constitute the most precise and flexible instrument of monetary management in the hands of the central bank. Federal Reserve purchase and sale of government securities in conjunction with, and to make effective, the discount rate has in theory, as well as in practice, overshadowed manipulation of legal reserve requirements as a tool of control.[1] In its continuing search for methods of improving the effectiveness of open market operations, the Open Market Committee appointed, in 1951, an *ad hoc* subcommittee to make a thorough study of the government securities market and the techniques employed by the System.

The principal recommendations of this subcommittee were adopted on March 5, 1953, following the submission of its report

1. This development has not been unchallenged. See, for example, the statement of John D. Clark in *Hearings*, Subcommittee on Economic Stabilization, 83d Congress, 2d session, Dec. 6–7, 1954, pp. 47–50.

to the Open Market Committee in November.[2] The proposals adopted were three in number: first, that open market operations henceforth be confined to the short end of the market; secondly, that support of the market during periods of Treasury refinancing be discontinued; and thirdly, that operations be conducted solely for the purpose of effectuating credit policy, and not for the purpose of supporting any pattern of prices and yields in the market.[3] The long-standing directive to the manager of the Open Market Account to "maintain orderly conditions in the Government securities market" was changed to one of "correcting disorderly conditions." While the latter phrase was not defined, it was interpreted to mean a qualification, in extreme circumstances, of the "bills only" doctrine contained in the first recommendation above.

It is submitted that these changes were not only of technique, or most significantly of method. Rather, they reflect a view of the responsibility of the central bank in the realm of economic stabilization — a view which differs in some respects from that of many monetary theorists.

The epitome of the thinking of the *ad hoc* subcommittee which resulted in the adoption of the new techniques was the belief that yields of government securities (and therefore the whole structure of interest rates) should be allowed to reflect the supply and demand forces of a free market. It believed that the free play of such forces would improve the market for government securities. Replies by dealers in government securities to a questionnaire sent to them by the subcommittee had strongly indicated the need for improvement and the subcommittee appears to have been influenced by these opinions.[4] Its report contains little analysis, however, of the effectiveness of the new techniques from the standpoint of Federal Reserve responsibility for economic stability. It appears to be the belief of the proponents of the new techniques that they will not only serve to improve the government securities market, but also that they will be adequate to the task of monetary control.

2. The full text of this report is published in *Hearings, ibid.,* pp. 257–307.

3. Board of Governors of the Federal Reserve System, *Annual Report,* 1953, p. 88. These "ground rules" were rescinded by the Open Market Committee on June 11, 1953, on motion of Mr. Sproul, only to be reinstated at the next meeting held Sept. 24, 1953. Members Sproul and Powell were alone in opposing the latter action.

4. For *ad hoc* subcommittee analysis of these replies see *Hearings, op. cit.,* pp. 292–99.

The purposes of this paper are (1) to examine and evaluate the underlying rationale of the decisions to confine operations to the short-term market[5] and to minimize System intervention in the market; and (2) to examine the effects of these decisions on (a) the ability of the Federal Reserve to contribute to economic stability and (b) the structural characteristics of the market. In general, I contend that the new techniques imply increased passivity in monetary management; specifically, I argue that they do not automatically improve the market's behavior or improve the effectiveness of open market operations.

II. Basic Arguments for a Free Government Securities Market

From the Treasury-Federal Reserve "accord" in March 1951, to the decision two years later which confined operations to the short maturity sector, Account transactions were conducted not only in bills, but in intermediate and long-term obligations as well. The following table[6] indicates the relative importance of operations in the various sectors of the market.

Open Market Account Transactions in U. S. Government Securities[1]
July 1, 1951–September 30, 1952
(In millions of dollars)

Class of security	Total		During periods of refunding		Other than periods of refunding	
	Purchases	Sales	Purchases	Sales	Purchases	Sales
Maturing issues (rights)..	3,059	....	3,059	....	....	....
Other securities maturing:						
Within 91 days.......	1,568	2,206	541	372	1,027	1,834
91 days–14 mo........	594	2,277	341	1,154	253	1,123
14 mo.–5 years.......	1	....	....	....	1	....
5–10 years..........	3	....	....	....	3	....
Over ten years.......	23	5	6	3	17	2
Total.............	5,248	4,488	3,947	1,529	1,301	2,959

[1]Excludes repurchase agreements with dealers and brokers and purchases and sales of special certificates from and to Treasury.

5. While the directive of March 5, 1953 referred to limitation of operations to the short end of the market, in practice the System has operated only in the bill market.

6. Taken from *Hearings, op. cit.,* p. 265.

The *ad hoc* subcommittee concludes that the "$32 million of trans-
actions in the intermediate and long-term sector are the only ones
that could properly be described as undertaken by the Committee to
'maintain an orderly market' "[7] in this period. This statement under-
estimates the extent of support in two respects. First, repurchase
agreements, which are excluded from the table, were frequently made
with dealers for this purpose. When the Federal Reserve Banks were
authorized to enter into such agreements at the Open Market Com-
mittee meeting on October 4, 1951, the Committee felt that "the
authority . . . should be available to each. . . . Bank so that it could
be used in the interest of orderly conditions in the Government
securities market."[8] Secondly, it would be justifiable to include as
support operations the $3,059 millions representing purchase of rights
to maturing issues. These purchases were often undertaken during
periods of Treasury refinancing to support the market.[9] In spite of
these qualifications, the evidence indicates that the System was
moving in the direction of heavier reliance on the bill market in the
two years following the "accord."

It was in this setting that the *ad hoc* subcommittee approached
its task and made its recommendations to the full Open Market
Committee. On the assumption that the effective execution of opera-
tions to control bank reserves depends upon the existence of "depth,
breadth and resiliency" in the market, it was concerned with the
"thinness" which had characterized the long-term sector after the
abandonment of the pegging technique in March 1951. This condi-
tion was ascribed to the uncertainty of dealers as to the sector of the
market in which operations would be conducted. To the subcom-
mittee the remedy was apparent: confine intervention in the market
to that necessary to supply and withdraw bank reserves; operate
exclusively in bills; and assure the market that these "ground rules"
would henceforth prevail.

Subcommittee concern with the structural characteristics of the
government securities market is understandable. Since open market
operations are the medium through which a significant part of
Federal Reserve credit policy is effected, success depends to some
extent on the adequacy of the market. Given the large volume of

---

7. *Loc. cit.*
8. Board of Governors, *Annual Report* (1951), p. 108.
9. The subcommittee gives recognition to this fact in a later section of its
report. See *Hearings, op. cit.*, p. 269.

System transactions, the desirability of a deep, broad and resilient market is not subject to question. Such a market would be characterized by (1) orders to buy and sell both above and below the current market price, (2) a large volume of such orders, and (3) relatively small fluctuations in price due to quick response on the part of investors to small changes in market conditions.

It was the belief of the subcommittee that these characteristics could be developed only in a free market. The following excerpts from the subcommittee's report indicate the strength of this contention:

". . . dealers must be confident that a really free market exists in fact, i.e., that the Federal Open Market Committee will permit prices to equal [sic] demand and supply without direct intervention other than such as would normally be made to absorb and release reserve funds."[1]

". . . it is not enough for the development of an adequate market that the Committee's intervention be held to a strict minimum. It is important that the dealers be assured . . . that the Committee is prepared to permit a really free market . . . to develop."[2]

". . . the guidance of economic decisions by free markets is a characteristic that has effectively served the American economy and for which there is no satisfactory substitute."[3]

These declarations were made with the full realization that "rates for Government securities are closely related to and affect interest rates on all classes of loans and investments."[4]

In commenting on the import of the new directives, Chairman Martin declared they "gave notice that the . . . Committee would not intervene to prevent fluctuations of prices and yields such as normally . . . occur as markets seek to establish equilibrium between supply and demand factors and to allocate savings. . . ."[5] It was his conviction, furthermore, "that we do the most service . . . consonant with the concept of private competitive enterprise, by giving the play of the market the maximum influence that it can have without disruptive effects."[6]

1. *Ibid.*, p. 266.
2. *Ibid.*, p. 267.
3. *Ibid.*, p. 300.
4. *Ibid.*, p. 301. For an interesting discussion of the relationship between the Treasury bill rate and the customer loan rate see D. A. Alhadeff, "Monetary Policy and the Treasury Bill Market," *American Economic Review*, June 1952, pp. 326–46.
5. *Hearings, op. cit.*, p. 23.
6. *Ibid.*, p. 229.

The free market thesis was not unqualified; the System stood ready to correct disorderly conditions should they develop. This possibility was considered "remote."[7] Furthermore, "the System would be called upon rarely, if ever, to intervene in securities with longer than 1-year maturity and . . . the only justification for System intervention would be to correct disorderly conditions resulting from some emergency, such as an unexpected development in international relations."[8]

Further evidence of the commitment to a free market is implied from the subcommittee's repeated assertion that the Committee should restrict its intervention in the market. In general, this meant that operations should be conducted for the single purpose of regulating the reserve position of the banking system. In formulating its report, the subcommittee declared that "the present wording of the directive of the . . . Committee on 'maintenance of orderly conditions' carries with it an unduly, *and even dangerously strong*, implication of continuing intervention in all sectors of the market. This prospect . . . seriously impairs the ability of the market to stand on its own feet or to evaluate correctly the real forces of demand and supply in the economy."[9] Testifying in December 1954, Chairman Martin stated that open market operations "have come to be limited to providing and withdrawing reserve funds. . . ."[1]

The arguments which the Federal Reserve authorities[2] muster in support of a free government securities market are of two types, specific and general. The specific arguments refer to the contribution which improvement of the technical operation of the market will make to the effectiveness of credit policy. The second type is quite broad: that yields and prices of government obligations *should* be determined in a free market. Although these arguments are interdependent, clarity of analysis dictates postponement of the discussion of the former to a later section. (IV below.)

Analysis of the statements of Reserve officials reveals the pre-

7. *Ibid.*, p. 304.
8. *Ibid.*, p. 298. The fact that long-term federal bonds "broke 90" without System intervention shortly after the new directives were adopted indicates the strength of its conviction in this respect.
9. *Ibid.*, p. 268, italics mine.
1. *Ibid.*, p. 21.
2. Not all members of the Committee agree with all aspects of the present policy. Mr. Sproul has sharply criticized the extreme free market position of the majority, as well as the manner in which it has been implemented.

sumption that the benefits of free, competitive markets for goods accrue to a free market for government securities. It is implied that the yields of government securities, like other competitively determined prices, serve as an allocating device. If the forces of private supply and demand are allowed free play in the securities market, the resultant yields serve to allocate funds among the several maturities according to the preferences of market participants. Furthermore, the level of interest rates on such securities in general serves to *measure* the impact of basic forces at work in the economy. As a prominent dealer argues, only in a free market "can the fluctuations in interest rates reflect accurately the supply and demand of Treasury securities that constitute the country's response to changes in business activity. . . ."[3]

Federal Reserve intervention in large volume in all sectors of the market, it is argued, inhibits the performance of functions assigned to the interest rate. First, it may result in a less-than-optimal allocation of savings from the standpoint of individual and institutional participants in the market. Specifically, intervention may require a rearrangement of funds, not upon the basis of private expectations, but on the basis of "artificial" conditions induced by System action.

Secondly, if the interest rate level is subject to official discretion, it is contended that the tightness or ease of the money market cannot be measured. Indeed, the effect of supplying or absorbing reserves on the banking system is obscured if interest rates are not allowed to *reflect* System policy. In this respect, the rate of interest is considered the prime indicator of credit conditions.

Before turning to the evaluation of these arguments, it is worth noting two alternatives to a free market which were considered by the Reserve. The first, involving rigid pegging of prices and yields of government securities, was rejected on grounds that are familiar to students of postwar monetary policy. The second, which was officially (if not in fact) the policy following the "accord," is generally considered commensurate with flexible monetary management. It involves the following guides to action: first, prices and yields are normally allowed to fluctuate freely as long as these fluctuations are not considered detrimental to the maintenance of full employment and price stability. Secondly, in the event that such fluctuations are deemed to be inconsistent with System objectives, the Open Market Committee would intervene in any sector of the market for government securities.

3. *Hearings, op. cit.,* p. 109.

This alternative involves Federal Reserve responsibility for over-all supervision of prices and yields of government securities — and in a sense, for interest rates in general. It does not involve continuous and inflexible control.

According to Mr. Martin, this variant was rejected because "it did not appear to offer real promise of removing obstacles to improvement in the technical behavior of the market."[4] It is worthy of note that the basis for this decision was not the relative merit of the alternatives with respect to the achievement of economic stability and growth.[5]

III. Evaluation of the Free Market Hypotheses

In modern employment theory the rate of interest plays a significant, if perhaps secondary, role. In Keynesian terms, the market rate relative to the marginal efficiency of capital influences the level of investment. Although the latter is usually the more volatile, changes in the market rate are not without importance. This is especially true of long-term investment. Given the marginal efficiency of capital, the higher the long-term rate the lower will be the volume of real investment, and conversely.[6]

As one of the strategic elements in the employment system, the interest rate has been the subject of policy prescriptions. In the main, these involve central bank supervision of long-term yields through open market operations. A concluding paragraph in Keynes' *A Treatise on Money*, written at the beginning of the Great Depression, declared:

"The remedy . . . [for instability] should come, I suggest, from a general recognition that the rate of investment need not be beyond our control, if we are prepared to use our banking systems to effect a proper adjustment of the market-rate of interest."[7]

The fact that this statement reflects a great deal more faith in the impact of the rate of interest on investment than the later Keynes (and Keynesians) would manifest is not the point. The point is that relatively few monetary theorists today would look with equanimity

4. *Ibid.*, p. 23.
5. There is no examination of this problem in the subcommittee report, and official statements from the Open Market Committee have ignored it.
6. K. K. Kurihara, *Monetary Theory and Public Policy* (New York: Norton, 1950), p. 119.
7. Vol. II, p. 386.

on central bank policy based on freely determined interest rates. Professor Chandler seems to state a predominant view:

"There are some respects in which the money market should be free of continuing official control. . . . But to allow the total supply of money and loans, and the price of loans, to be determined by private demand and . . . supply would negate the very idea of central banking. . . . It should be clear that the Federal Reserve can make its maximum contribution to economic stability and growth only by recognizing its continuous responsibility for money market conditions. . . . A successful policy of economic stabilization cannot be a passive policy."[8]

With respect to System responsibility for the rate of interest, Chandler argues as follows:

"It also needs to be emphasized that a shift to the objective of promoting economic stability . . . does not mean that the Federal Reserve should cease to be concerned . . . [with] interest rates, nor that its control of interest rates should be any less precise than was its control during the pegging period. . . . [Its] mistake during the pegging period was not that it controlled . . . rates; the mistake was in stabilizing interest rates — in making stability . . . an overriding objective and in sacrificing all other objectives."[9]

The Federal Reserve assumption that free product markets are analogous (in a welfare sense) to a free market for government obligations is subject to question. Some sense may be made of a policy which allows supply and demand to allocate funds among the various sectors of the market. Beyond this, to allow the *level* of rates to reflect free market forces is to underestimate the importance of the interest rate as a price. There are a number of reasons why interest rates on government securities cannot be viewed in terms of competitive price analysis.[1] Since an exposition of this kind is not within the scope of this paper, the following remarks are merely illustrative of the proposition.

Considering first the impact of a price increase (for example) in the market for a consumer good and a similar change in the yield of a government security, it is obvious that the significance of the latter is relatively greater. While the list of repercussions which follow the change in price of the consumer good may be long, it is probable that the resulting reallocation of resources (in spite of short-run immobilities) will leave total national employment little changed.

8. *Hearings, op. cit.*, p. 46.
9. *Loc. cit.*
1. Assuming, of course, that the goal of monetary policy is economic stability at near-full employment.

This reallocation follows consumers' preferences under competitive conditions, and whatever frictional unemployment results is considered the normal price paid for free choice.

On the other hand, a rise in the yield of government bonds, for example, has a greater impact on the economy because of the very size of the federal debt. The yield on other debt instruments of similar maturity will tend to rise. The effect of this on some kinds of investment is likely to be great. Recent theoretical and empirical studies indicate that demand for investment in public utilities (and outlays of a similar nature) has considerable interest elasticity.[2] In the postwar years 1946–51, total expenditure for new plant and equipment by transportation, electric, and gas utilities was approximately 23 per cent of total business investment in new plant and equipment for this period.[3] Given the magnitude of such investment, a rise in the bond yield may reduce employment substantially — particularly so if the increase in yield is due to a shift in liquidity preference and there is no causal or independent increase in the propensity to consume.

A second major impact of a rise in yields arises because a significant share of the community's saving is placed in this market through institutions. A gradual rise in yields (which does not call for System intervention, since it does not imply disorderliness) would reduce significantly the capital value of these savings, as well as those which are placed in nongovernmental bonds. While it is true that a rearrangement of portfolios would ameliorate the adverse effects of the yield change, it is unlikely that this could be accomplished without substantial losses. This result seems paradoxical in light of the System's present intention to influence as little as possible the allocation of funds in the market. Passivity in this case is at least as disruptive as positive intervention.

A third effect of the rise in yields is the resultant increase in the burden of the public debt. Rising yields require higher coupon rates on new issues if the Treasury is to prevent attrition to its balance.[4] This in turn increases annual service charges on the outstanding debt. If national income is rising, the increase in the burden may be

2. Cf. R. S. Sayers, *Modern Banking*, 3d ed., p. 197; and L. W. Mints, *Monetary Policy for a Competitive Society*, pp. 201–2.

3. Derived from *The Mid-year Economic Report of the President*, 1952, Table B-18, p. 157.

4. This is particularly true since the decision to withdraw support of the market during Treasury refinancing.

negligible.[5] Even if this is the case, a transfer burden may remain. Recent studies, admittedly based upon debatable assumptions, indicate that the transfer burden of the debt is not so great as commonly believed.[6] In spite of these conclusions, a greater burden may result in the form of decreased current government welfare services if Congress is disposed to maintain the budget at the level existing prior to the increase in rates.

These considerations emphasize the economy-wide importance of government security prices and yields. If the government debt were a small fraction of total debt, arguments for a market of the type envisaged by the present Open Market Committee (and many dealers) would assume greater validity. While the objective of Federal Reserve policy remains that of economic stabilization, its insistence upon a free market appears to lessen, rather than enhance, its powers in this respect.

This conclusion seems justified in light of the subcommittee's declaration that most non-self-correcting movements in prices and yields of government securities "reflect basic changes in the credit outlook and should not be the occasion for intervention"[7] by the Open Market Committee. It is precisely these "basic changes" that central banks, including the Federal Reserve System, have in the past sought to control; indeed, their *raison d'etre* lies in the necessity for controlling credit movements of this kind.

In some instances, the free movement of yields and prices of government securities might coincide with the objectives of monetary policy. For example, an increase in the level of yields in the government securities market when inflationary pressures exist in a fully employed economy would not be deemed undesirable. Nor would falling yields be inconsistent with monetary policy during periods of falling employment. In both cases, the change in yields will have *some* desirable effect on that part of total investment which is interest-elastic.

Even if the above is granted, two objections to the present Federal Reserve position remain. In the first place, it is conceivable that the *rate* of change of government securities' yields will not coincide with the desired rate. Even though the direction of change

5. Cf. E. Domar, "Public Debt and National Income," *Public Finance and Full Employment*, Postwar Economic Studies No. 3, Federal Reserve System, pp. 58–60.

6. D. C. Miller, *Taxes, The Public Debt, and Transfers of Income*, pp. 141–42.

7. *Hearings, op. cit.*, p. 268.

in yield might be desirable, it is extremely unlikely that free-market decisions will always result in an appropriate rate of increase or decrease. The rapid increase in the short- and long-term yields in late 1952 and early 1953,[8] during a period characterized by a falling wholesale price level, consumer price stability, and rising industrial output, may have exerted an inordinate dampening effect on investment decisions. It seems reasonable to suppose that Federal Reserve action both in the long and short end of the market (or even the latter alone) to control the rate of increase in yields would have been desirable at that time.

In the second place, there is reason to believe that yields in the various maturity sectors of the government securities market might, under certain circumstances, move in opposite directions. For example, if the short-term rate is falling in a recession, this desirable movement may not be accompanied by a similar change in the bond rate. This possibility, which is discussed in more detail in a later section, would reflect a change in "basic credit conditions" and, as well, a change in the liquidity requirements of market participants. If the statements of Reserve officials are to be taken literally, this would not call for System intervention. Intervention, however, is clearly desirable. Passivity on the part of the central bank in this case can only mean continued postponement of marginal long-term investment necessary to recovery.

While Professor Harris' statement that modern "developments in the theory of money and output seem largely to have escaped those responsible for monetary and debt policy"[9] is perhaps too strong, his analysis of the viewpoint of present Reserve officials is essentially correct:

"They seem to consider the control of the rate of interest on Government securities merely as an attempt to raise artificially the price of these assets rather than (as they should) consider the control on this rate as a means of determining the rate of interest generally and hence influencing investment and output. . . ."[1]

If the above conclusions are reasonable, are there any compensating advantages in a free government securities market? This

8. Yields on Treasury bills rose from 1.700 in June 1952 to 2.231 in June 1953; yields on long-term bonds increased from 2.61 to 3.09 in the same period. *Hearings, op. cit.,* p. 29.

9. *Ibid.,* p. 55.

1. *Loc. cit.*

question is given a strong affirmative answer by a majority of the Open Market Committee.

IV. Technical Advantages of a Free Market

A free market, according to its proponents in the System, promotes market depth, breadth and resiliency. These characteristics, it is contended, facilitate the Committee's task of supplying and withdrawing reserves from the banking system. The problem of "thin" markets following the "accord" was a result of (1) dealer hesitation to take positions and engage in arbitrage when the threat of System intervention was omnipresent, and (2) uncertainty on the part of dealers and other investors as to the meaning of the directive to "maintain orderly Government securities markets."

The *ad hoc* subcommittee argued that dealers should play a large role in making a sound market. In order to be in a position to perform this function, they should not only provide brokerage services, but also be active in buying and selling for their own accounts. The latter involves willingness to take position risks and to engage in arbitrage. By questionnaire, the subcommittee found that dealers were largely confining operations to simple brokerage.[2] This was found to be a result of the propensity of the Open Market Committee to execute its operations in all sectors of the market. Dealers complained that difficulties inherent in predicting the turns of the market were compounded by the necessity of forecasting the maturity in which the Account manager would choose to conduct purchases or sales.[3]

The contribution of arbitrage operations to a smoothly functioning market are too well known to require detailed treatment at this point. The following illustration, however, will clarify the subcommittee's concern.

Assume, for example, a dealer who feels the existing relationship of prices and yields between the short- and longer-term government securities does not accurately reflect the real or prospective forces in the market.[4] His analysis will (in the absence of risk of System intervention) result in recognition of opportunity for profitable opera-

---

2. *Hearings, op. cit.,* p. 21.
3. *Ibid.,* p. 267.
4. As a general term, "arbitrage" may be used to describe dealer reaction to this situation. In a technical sense, however, it usually refers to dealer operations in two securities which are closely related in maturity.

tions on his own account utilizing his own and borrowed funds. To make the illustration concrete, assume that he believes the spread between prices in the two sectors is abnormally small. On the strength of intuitive estimation he will sell the short and purchase the longer maturity. If his expectations are borne out through changes in market forces, he will make a profit. The effect of taking position risks of this nature, where it is continuous, is to reduce price fluctuations, to narrow the spread between bid and ask quotations, and to provide a substantial volume of activity in all sectors.

Assume now that the System is conducting operations in all sectors of the market. If, after taking his position, open market sales are executed in the long-term maturity, the forces in the market anticipated by the dealer are obstructed. The risk of sustaining losses is "abnormally" increased by System intervention. Aware of this possibility, the dealer would be reluctant to take a firm position. "To do so would not only involve the risk of being wrong in . . . evaluation of economic and market trends, but also of being wrong in guessing at what point the . . . Committee might . . . intervene."[5] Assurances that the Committee would confine its operations to the bill market would, it was contended, eliminate this hesitancy. Further assurance that System operations would not be conducted to influence prices and yields would re-enforce the confidence of market participants.

The above position is further strengthened when one considers the structure of the government securities market. Dealers, of which there are six of major significance, affect to a large degree the functioning of the market. They have extensive knowledge of supply and demand at any given moment in their capacity as brokers for a large segment of institutional investors. Furthermore, they may be in the market on their own account. The funds with which they operate in this latter respect are largely borrowed. The ratio of borrowed money to dealer capital may be as high as 100:1.[6] To a dealer very small changes in prices and yields are of great significance; a one/thirty-second change may be the difference between profit and loss.

Dealers contend that they are willing to take the risks of a free

---

5. *Hearings, op. cit.*, p. 267.
6. The source of these funds may be the New York Banks or free reserves in other parts of the country. The volume of dealer participation at the present time is approximately $1 billion; the interest paid by dealers for use of these reserves is usually the federal fund rate.

market. On the other hand, they are naturally hesitant to take positions involving large sums of borrowed money in a market characterized by more or less continuous intervention in all sectors.

These considerations lead to the conclusion that "bills only" operations and minimum intervention *should* improve the depth, breadth and resiliency of the market.[7] Acceptance of this generalization does not, however, justify the conclusion that the present techniques will be effective instruments of monetary management for achieving economic stability. The *ad hoc* subcommittee clearly made a conclusion of this nature. Since there is no analysis of the problem in its report, it must be inferred that the subcommittee believed the present techniques adequate to the task. It is submitted, however, that there are limitations and defects with respect to the present techniques which require examination.

V. "BILLS ONLY," MINIMUM INTERVENTION, AND MONETARY MANAGEMENT

Analysis of the effectiveness of present open market practices involves discussion of three interrelated questions. First, to what extent may it be expected that operations confined to the bill market will result in desirable changes in the intermediate and long-term sectors? Second, is it possible that a bills-only policy may conflict with the intention of the System to intervene as little as possible in the market in supplying and absorbing reserves? In other words, may there not be times when operations in the longer maturities will give quicker results with less disturbance to the market than bills-only operations? Third, are there any inherent limitations to the effectiveness of a bills-only policy?

A. *Effects of Operations in the Bill Market on the Long-Term Sector*

It was the subcommittee's view that most market fluctuations are either (1) self-correcting or (2) reflections of basic changes in credit outlook which are not self-correcting. In neither case is intervention considered desirable or necessary.[8] The possibility was

---

7. Sproul believes the new techniques have not resulted in an improvement of the market. "We must not be misled by the claims of . . . dealers who urge the present techniques and now proclaim that they are helping to create a broader market." *Hearings, op. cit.*, p. 226.

8. *Ibid.*, p. 268.

admitted, however, of a change in long-term prices and yields which would not belong in either of the categories above. It is with this latter situation that we are immediately concerned.

Assuming, for example, continued selling pressure in the bond market which did not appear to be self-correcting, i.e., resulting in counteracting arbitrage, what action would be required of the Reserve Open Market Committee? According to the subcommittee, the great majority of such cases "will respond readily to arbitrage induced by positive intervention on the part of the Committee in the very short sector of the market."[9] Only rarely, if ever, would the situation become so disorderly as to require intervention directly in the long maturity.

The reasoning of the subcommittee seems to be as follows: A persistent downward trend in the price of bonds (and a concurrent rise in yields) will be arrested by System purchases in the short-term market. This action, by the nature of the case, must be pursued to the point where the price of bills is increased and their yield reduced. As soon as this has been accomplished, *induced* arbitrage operations will occur. The higher the bill price rises, the more will opportunity exist for profit; dealers are expected to sell short-maturities (bills in this situation) and buy the (now) undervalued bonds. In this manner funds are assumed to flow smoothly and quickly from the short end of the government securities market to the long-term sector. The increase in demand for bonds will arrest the undesirable movement in that part of the market, without the necessity of direct intervention by the Committee.

The unqualified optimism of the subcommittee with respect to the impact of bill rate changes on the long-term market is not shared by many authorities. Riefler, while impressed with the tendency of short- and long-term rates to move together in the same direction, observed that "funds do not always flow freely between the two markets simply in response to rate changes."[1] Professor Adarker, who extended Riefler's statistics to cover the period 1929–32, found that the tendency to equi-proportional variability of long- and short-term yields which characterized the 1920's did not project into the early depression years. And this occurred in spite of an artificial

9. *Loc. cit.*
1. W. W. Riefler, *Money Rates and Money Markets in the United States*, pp. 122–23.

lowering of short-term yields.[2]   Keynes, in the *Treatise*, gives ample rationale for a sensitive response of the long-term rate to short-term variations.[3]   He contends that, under normal conditions, the central bank's manipulation of the short-term rate will cause desired changes in bond yields.   Yet in the final analysis, he questions the ability of the central bank to make its policy effective when "prices are falling, profits low, the future uncertain and financial sentiment depressed. . . ."[4]   In other words, when a slump is developing, even short-term purchases *à outrance* may not be sufficient to bring down the long-term yield.   This is so because "it is precisely at such a time . . . that lenders are exigent and least inclined to embark their resources on long-term . . . so that the bond rate, far from falling . . . may be expected . . . to be higher than normal."[5]   In these circumstances, what can be done other than to *"impose on the Central Bank the duty of purchasing bonds up to a price far beyond what it considers to be the long-period norm"*?[6]

In a recession, then, there are a priori grounds for believing that bill operations will be ineffective in arresting weakness in the longer maturities.   Against the positive effect of dealer arbitrage, one must consider the negative effect of investor expectations as to the future level of the long-term rate.   Suppose the majority of holders of a recently issued 2 per cent, 30-year bond (and non-owners as well) believe the normal return on such investments will rise to 3 per cent two years hence and stay at this level indefinitely.   Heavy selling will ensue; with these expectations, no investor would be willing to pay much more than $740 for the bond, the par value of which is $1000.   The approximate loss of any seller in such a situation would be the original purchase price minus the price existing at the time he sells, plus any interest lost from the time of purchase.   But this would be less than the loss of holding onto the bond below a price of approximately $960 — depending upon how quickly the market is declining.[7]

2. B. P. Adarkar, *The Theory of Monetary Policy*, pp. 113–18.   It is significant that Adarkar believes the traditional causal nexus between the two sectors is in error, i.e., that short-term movements do not *cause* changes in the long sector.

3. *Op. cit.*, pp. 356–62.

4. *Ibid.*, p. 373.   This situation would not seem to fall within the System definition of disorderliness.   Indeed, since it constitutes a "change in credit outlook" it might not call for intervention at all.

5. *Loc. cit.*

6. *Loc. cit.*   The emphasis is Keynes's.

7. If he had sold early, he could have placed his funds in the short-term market at a low interest rate.   The point of "indifference" depends partly on the latter rate.

In this case, the long-term rate may be expected to move in a direction opposite to that of the short-term yield. If the central bank pursues its bills only policy, its efforts will be futile if the forces of market expectations outweigh the influence of dealer action. Heavy Federal Reserve support of the bill market, involving purchases and declining bill yield, may be reinforced by private demand for short-term securities as the public moves out of the bond market. Thus, "public opinion concerning the future course of interest rates is a very powerful factor in the long-term market, and a mere reduction of short-term rates may not do much to modify that opinion; and unless that opinion is modified little impression will be made upon long-term rates even if the authorities are prepared to operate in the market on a very large scale.[8]

The historical evidence concerning the effectiveness of bills-only operations in such situations is not only scanty but mixed. On the optimistic side, the Federal Reserve may call attention to the rise of bond prices after its shift to active ease in early June 1953. Large-scale purchases were made in the bill market after bond prices had dipped to extremely low levels. There is evidence that dealers covered their shorts with rather heavy purchases of bonds both after and a few days before the System intervened.[9]

On the other hand, there have been a number of times when central bank operations in the short end of the market have not been effective in bringing about desirable changes in yields of long-term securities. In 1932, for example, British monetary authorities were successful in lowering the short-term rate considerably through large purchases. The long-term rate, however, remained high.[1] This is a classic illustration of the fact that funds may not flow (as the Federal Reserve officials contend they will) from the short end of the market to the long.

It would seem, therefore, that the success or failure of bills operations in arresting undesirable movements rests upon the sensitivity of public reaction to them. The optimism of present Reserve officials with respect to the efficacy of such operations is not entirely justified. Both theory and a part of the empirical evidence would lead to a qualification of this state of mind.

8. Sayers, *op. cit.*, p. 188.
9. Mr. Sproul believes that the Committee, in taking credit for this success, has embraced the *post hoc, propter hoc* fallacy.
1. Sayers, *op. cit.*, p. 189.

Finally, the Reserve position that movements in the market resulting from "changes in credit outlook" do not call for intervention in any sector is subject to criticism. Although this has been covered in a previous section, it should be emphasized again that the job of the central bank is essentially one of control. To restrict this control to the level of bank reserves is to discard the advantages of direct supervision of the interest rate and (at times) the volume of investment which is sensitive to this variable.

B. Internal Consistency of the Bills-Only Doctrine with Minimum Intervention

One of the principal arguments for a bills-only policy is that it is consistent with the objective of restricting intervention in the market to absorption and provision of bank reserves. The Open Market Committee accepts the view "that a more self-reliant market in . . . Government securities would develop if its intervention were solely to effectuate the objectives of monetary and credit policy and were carried out by making purchases and sales in the short end of the market."[2]

Under normal conditions it is probably true that bill operations will have the least impact on the market. System purchases and sales in this sector least affect the asset value of investor portfolios.

An inconsistency may arise, however, between bill operations and the ideal of minimum intervention. Assuming again a non-self-correcting weakness in the price of bonds, will bill market intervention necessary to induce arbitrage be greater or less than would be required through direct intervention in the long-term sector? No general answer seems possible. If the weakness is due to a change in liquidity preference or other market expectations, it is quite possible that intervention in bills will have to be heavy and sustained.[3] Mr. Sproul contends "that on some occasions the System might better . . . effectuate its policies by operating in other sectors of the market — even the longest maturities — depending on the economic conditions then prevailing. . . . Restriction of operations to short-term securities would probably either make the System's intervention ineffective or require larger scale intervention to achieve the objectives."[4]

2. Board of Governors, *Annual Report*, 1953, p. 89.
3. Keynes, *op. cit.*, pp. 370 ff., and Adarker, *op. cit.*, pp. 117–18.
4. *Hearings, op. cit.*, p. 311. One government securities dealer contends that open market operations confined to bills after June 1953 caused a disconcerting disruption of the normal functioning of this market.

There is some reason to believe that bill intervention will have to be heavy if the central bank has long postponed action. Furthermore, market expectations will respond more readily to direct intervention in the longer maturities than to bill operations. This intervention need not be continuous or *à outrance*. If the market expects intervention in the long-term sector in such cases, its normal expectations will tend to be revised quickly in response to moderate operations. This argues for the "middle of the road" policy which was rejected by the Committee on the ground that it would not improve the characteristics of the market.

C. *Implicit Limitations of Bill Operations*

These considerations cast doubt on the ability of the central bank to maintain effective monetary control within the framework of a bills-only policy. Further examination of the problem reveals the existence of still other limitations.

The first might appear in connection with attempts to reverse an incipient downward movement of *rates* on long-term securities which is not self-correcting. Inducement of arbitrage through bill sales might be limited by the lack of sufficient bills in the central bank's portfolio. If the upward movement of bond prices continued over a long period of time, the System might be forced to extend its operations to longer maturities. At the present time, for example, the System's holding of intermediate length securities is many times larger than its bill portfolio. In order to pursue its present course of dealing only in the short end of the market effectively, this distribution should be partially altered. This, however, would probably call for further shortening of the public debt — a policy that is not likely to find favor in the present Treasury.

Where the System is purchasing bills to effect credit ease, the same limitation appears in different form. Much of the supply of bills outside the Reserve is in the hands of holders who are not sensitive to changes in the price of these instruments. To these investors (corporations and some banks, for example) bills are considered a superior alternative to holding cash, and riskless in the sense that they mature to par. The small gain which a corporation would realize through a sale prior to maturity is generally not considered worth the trouble involved. Thus, a large segment of the bill supply is frozen; in order to dislodge sufficient sales to the System, it may be necessary to bid *higher* prices than would be required otherwise.

Finally, it may be observed that inducement of arbitrage is limited if the bill rate is already close to zero. Even very heavy purchases in the bill market alone may not be sufficient to halt a decline in bond prices.

VI. Conclusions

Few economists today would advocate exclusive reliance on central bank action to stabilize either prices or employment. In recent years the limitations of monetary tools have been emphasized and fiscal policy for stabilization purposes has assumed a predominant role. In spite of this development, it is nearly always assumed that the central bank *should* pursue policies which contribute to the reduction of fluctuations in co-ordination with those of the fiscal authority.

The current "experiment" in which the Open Market Committee is engaged does not appear to implement its power to promote economic stability. A major conclusion of this paper is that techniques which improve the characteristics of the government securities market are not necessarily those which contribute to the effectiveness of monetary policy. Indeed, it is highly probable that a free market objective lessens the degree to which the System may influence the level of employment.

This is particularly apparent when one considers the extent to which the Open Market Committee majority has embraced a laissez-faire policy with respect to interest rates.[5] The view that the interest rate is a thermometer which measures the health of the economy finds little support in recent monetary theory or best central banking practice. While the latter has traditionally involved operations in the short end of the market,[6] the *aim* has been, at least in recent years, to maintain some degree of supervision over the market rate of interest.

There is, furthermore, some doubt concerning the effectiveness of bills operations to induce arbitrage, the object of which is the correction of movements in the long-term rate. Ample theoretical grounds exist for a convergent movement of prices and divergent movement of rates in spite of heavy bill operations.

5. The recent statement of Chairman Martin to the effect that the System has not ceased to be concerned with interest rates (*Hearings, op. cit.,* p. 229) does not counterbalance a large number of official statements, including those of Mr. Martin, which imply the contrary.
6. The Bank of England has recently operated in long-term bonds on occasion, and with great success. See Sayers, *op. cit.,* p. 130 n.

It would seem reasonable to suppose that the current "ground rules" relating to bill operations and minimum intervention will give satisfactory results as long as the economy is essentially stable. Small fluctuations in prices and yields should neither cause concern nor result in direct intervention in the bond market. But one cannot agree that movements which are the result of changes in private credit outlook should be immune from System intervention. If the central bank is to influence the level of long-term investment with any precision, it is necessary that it be willing to supervise the bond rate. In some cases this may be accomplished by pumping reserves into the short end of the market. In others, it may be more effective (from the standpoint of volume and timing) to intervene directly in the bond market. While this may involve a loss of depth, breadth and resiliency in the market, there is some reason to believe that this is a small price to pay for greater monetary control.

<div align="right">Deane Carson.</div>

Brown University

INVESTMENT CRITERIA, PRODUCTIVITY, AND ECONOMIC DEVELOPMENT*

By WALTER GALENSON AND HARVEY LEIBENSTEIN

I. A suggested criterion for investment allocation, 343. — II. The case for high labor productivity in underdeveloped areas, 353; the level of labor productivity in underdeveloped areas, 353; labor productivity and capital allocation, 355; population growth and development, 363; some obstacles to high labor productivity in backward areas, 367. — III. Conclusion, 369.

I. A Suggested Criterion for Investment Allocation

In recent years there has been increased attention to the problem of establishing criteria for allocating investment in programs for economic development. A general rule has emerged, from which practical policies are inferred. A. E. Kahn, in an article that appeared in 1951, set forth the "rule of social marginal productivity" as a guide to investment, and deduced, among other things, that as a consequence of this "rule," underdeveloped areas should choose industries and techniques requiring a lower capital/labor ratio than that prevailing in developed countries.[1] Hollis B. Chenery accepts the criterion of social marginal productivity, and attempts to demonstrate its application to a number of empirical situations.[2]

We propose in this paper to examine the conclusions that have been reached, and to suggest a line of reasoning that appears to us to be more in consonance with the peculiar problems raised by economic development. We do not intend to advance a complete model for investment allocation; that would obviously be impossible within the confines of a brief paper. Our object is primarily to call attention to the shortcomings occasioned by the failure of economic theory in dealing with economic growth, to relax some of the assumptions that are relevant in treating static problems. We shall also indicate briefly some of the institutions and practices characteristic of back-

ward areas which we believe must enter into the framework of a theoretical formulation suitable as a guide for successful economic development.

Chenery voices the general proposition that "Economic theory tells us that an efficient allocation of investment resources is achieved by equating the social marginal productivity of capital in its various uses."[3] With this bald statement we can have no quarrel. However, it is open to the same general objection that Friedman raised with respect to Lerner's famous "Rule": standing by itself, it provides no guide to policy.[4] The real question arises when one attempts to make precise the notion of "efficient allocation" in this context.

In economic statics when we consider allocation of resources, we can attach a clear-cut test to the idea of efficiency. The test is the maximization of the value of the national product. That allocation which maximizes the value of national product is the efficient one — and equating marginal productivity in different uses is the rule for achieving such an allocation. But even here some ambiguity arises in the case of capital goods, and this may have something to do with the imperfect state in which capital theory finds itself. The difficulty arises because of the *ceteris paribus* assumption[5] necessary in comparative statics, e.g., we do not know of what value these capital goods will be to generations yet unborn. With respect to the production of consumer goods the situation appears to be fairly clear, since the valuation of the goods depends on their worth to the population in the current period, rendering the *ceteris paribus* assumption reasonable. But the valuation of capital goods depends on the value of the output stream generated by an increment of capital projected into the indefinite future. Apart from difficulties of predicting the output stream, there is also the problem of knowing what meaning to give to the *ceteris paribus* assumption in situations involving the indefinite future, as well as the practical consideration as to the value of the assumption under conditions in which technology, tastes, and population will most certainly change.

The question that immediately arises in applying conventional theory to problems of economic development is whether the goal of

3. Chenery, *op. cit.*, p. 76.

4. Milton Friedman, "Lerner on the Economics of Control," *Journal of Political Economy*, Oct. 1947, p. 405.

5. I.e., tastes, size and composition of the population, the state of the arts and expectations remain the same. Or, another way of looking at it is that an "efficient" allocation requires an accurate forecast of tastes, population, and the state of the arts in the future. See also Lionel Robbins, "On a Certain Ambiguity in the Conception of Stationary Equilibrium," *Economic Journal*, June 1930, pp. 159–79, esp. p. 168.

maximization of the national product necessarily leads to "development." Suppose that in every period 110 per cent of the GNP is consumed, would we then say that development is taking place even if in period after period the GNP is the maximum it could possibly be under the circumstances? As soon as we leave the theoretical world of statics, maximizing output is no longer a sufficient criterion, nor does "efficiency" have the same meaning.

To get at the correct criterion, we must determine the appropriate goal of the economy during the process of development. Unfortunately, this cannot be done on the basis of economic analysis alone; a social welfare function of some sort is necessary. The goal must depend ultimately on values that come from outside economic analysis, and these are arbitrary to some extent. Nevertheless, a good case can be made for the proposition that the appropriate economic goal should be the maximization of *per capita* output, or *average* income, either over time, or at some time in the future.[6] This is hardly a startling position: it appears to be quite widely accepted. But it should be noted that maximizing per capita income is not at all the same as maximizing the national product once the *ceteris paribus* assumption is dropped.

For a closed economy, *per capita* income and *per capita* output are identical, but there may be considerable deviation between *per capita* output and the *per capita* level of consumption. If we are interested in alleviating the mass poverty that prevails in backward areas, then ultimately we must be interested in their *per capita* level of consumption. However, in the short run raising *per capita* output and raising *per capita* consumption may be antithetical, for the rate of capital accumulation will depend on the extent to which increases in output are not followed by equal increases in consumption. Thus, while the raising of *per capita* consumption levels may be, and perhaps must be, an ultimate goal of development, if we concentrate on consumption as an immediate goal, it may be impossible of achievement as an ultimate goal. It is therefore *per capita* output that must be looked upon as an appropriate index of economic development.

We turn now to a consideration of the extent to which the practi-

6. It should be noted that there are numerous and difficult problems involving the element of time that we do not consider in this paper. For example, the ordering of income magnitudes over time is a matter fraught with most important considerations of economic welfare; and so is the precise future time at which it is desired to achieve maximum output. Unless the time path or the time horizon is given, there is no meaning to the concept of maximization in the future. As we shall see below, however, the time element is intimately related to specific variables, of which the rate of population change is perhaps the most important.

cal corollaries which have been deduced from the social marginal productivity rule (SMP) focus on this goal. Although the literature is far from clear on this point it seems to us that three general corollaries of SMP have been advanced as valid policy guides. They are: (1) to maximize the current output/investment ratio; (2) to maximize the labor/investment ratio; and (3) to maximize the export goods/investment ratio.

The discussion is often needlessly complicated by attempting to consider simultaneously development problems and balance-of-payments problems. Where investment funds or capital goods are obtained from abroad, a balance-of-payments problem *may* develop. The extent of the problem will depend, in part, on the allocation of the investment. But since the extent of the balance-of-payments problem will also depend on numerous other monetary conditions, both internal and external to the underdeveloped country, it is best, initially, to treat the two questions separately, and to leave the balance-of-payments problem for another occasion. We shall therefore omit corollary (3) in this paper, and consider only corollaries (1) and (2), and their relationship to the SMP criterion.

Kahn tells us that "The SMP of capital is not correlated with the rate of turnover."[7] The rate of turnover is of course the output/investment ratio. Yet later he concludes that ". . . general adherence to the capital-turnover criterion . . . is particularly desirable (a) where capital is relatively scarce and labor extremely plentiful. . . . With respect to the first circumstance (a), the rule of thumb of employing the minimum amount of capital necessary to absorb excess labor does indeed approximate the SMP criterion."[8] But there is no clear-cut explanation why this should be the case. There are, however, good reasons to believe that the rate-of-turnover criterion is not likely to be consistent with the proper interpretation of SMP from the point of view of development.

It is obvious that in computing social marginal productivity, we must consider the effect of an increment of capital on the output stream in the indefinite future, and not only on the magnitude of the output during the initial period. If the income stream is the same period after period, then the capital-turnover corollary would appear to favor, as Kahn correctly says in a footnote,[9] short-lived over long-

7. Kahn, *op. cit.*, p. 39.
8. *Ibid.*, p. 51.
9. The notion of short-lived vs. long-lived capital may need some amplification. For this purpose it may help if we reproduce the example given by Kahn (p. 39). "Compare the following financial data for investments *A* and *B*, which differ only in that equipment in *A* lasts only two years, in *B* ten years:

lived capital. But if national capital is to be maintained, and the total effect on the income stream is to be considered, then the length of life of any individual capital good is an irrelevant consideration. If the income stream increases over time, then the application of the capital-turnover criterion *can* quite clearly result in the allocation of investment to uses that have a lower SMP than the alternatives. For example, between two alternatives, *A* and *B*, the following might be the case: Equipment in *A* lasts five years while in *B* it lasts ten years. Assume that in both cases the income streams over the life of the equipment is twice the investment cost. Clearly the rate of capital-turnover is twice as great for *A* as for *B*. But if replacements are made so that capital is maintained intact indefinitely, then with a rising output stream for *B* and a constant output stream for *A* it is certainly possible that the SMP for *B* would be larger than the SMP for *A*. Where industries of increasing returns with respect to scale are among the possible alternatives, it is likely that the capital-turnover criterion will result in the wrong choice. A constant or decreasing returns industry may have a higher capital turnover ratio than an alternative increasing returns industry, but the increasing returns industry may in the long run make a greater contribution to national product.

If SMP is to be interpreted in this context to take account of the addition to the income stream in the indefinite future (no matter how discounted), then we cannot assume, as a static interpretation of SMP would have us, that other things remain constant during this process; we must take into account the time pattern of possible changes as a consequence of the addition to capital. Thus, with respect to many investment alternatives, we have to account for such dynamic factors as the development of skills, the development of markets, overcoming production bottlenecks, etc., and the effect of these factors on the time pattern of output — circumstances under which the capital turnover rule is likely to come up with the wrong answer. In the short run, the capital turnover ratio could be highest in those endeavors that required few new skills, that depended on

Investment...........	$100,000	$100,000
Annual Data:		
Output.............	60,000	20,000
Depreciation........	50,000	10,000
All other Costs......	20,000	10,000
Capital Turnover......	60 per cent	20 per cent

In a long range investment program, in which real national capital must presumably be kept intact, there would appear to be no advantage in choosing the shorter- over the longer-lived investment, since both yield the same annual output net of depreciation." However, as we will argue, there is an advantage in longer-lived investment because of the lower replacement cost in a growing economy.

existing markets, and in which there were no production bottlenecks
to overcome; but over the longer period it might be those investment
alternatives that involved the learning of new ways of doing things
that would make the greatest contribution to the social product.

We now turn to corollary (2), under which the labor/investment
ratio is to be maximized. The extent to which labor can be absorbed
in any economic process depends on the flexibility and adaptability
of the other factors of production (i.e., capital). One can easily
visualize situations in which the maximum labor absorption criterion
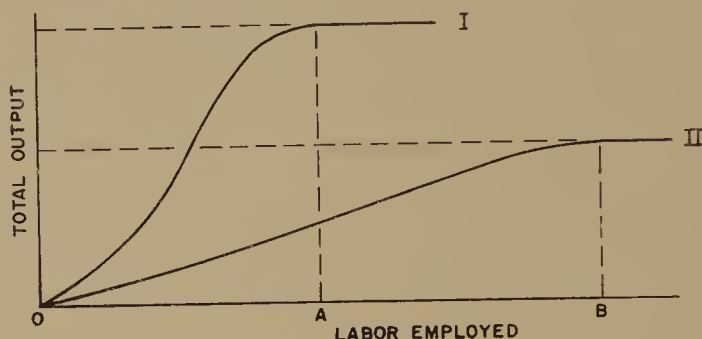would not maximize the addition to total output. Figure I illustrates

FIGURE I

such a situation. The investment in alternatives *I* and *II* are the
same. The maximum absorption of labor in either case is up to the
point where the marginal productivity of labor is zero, if we assume
that the marginal productivity of labor on the land is zero. In
Figure I, investment alternative *I* would absorb *less* labor but add
more to total output than would investment alternative *II*. As a
consequence alternative *I* rather than *II* is consistent with SMP.
Nor is there any reason to believe that highly productive capital is
necessarily related to highly flexible and adaptable capital. Indeed,
there is good reason to believe that quite the reverse is true. With
respect to machinery it is likely that the more productive the machine,
the more specialized it is, and as a consequence the less adaptable and
flexible with respect to changes in other factors.

A second set of difficulties has to do with the time element.
Corollary (2) does not tell us in any clear way when the labor is to be
absorbed. If the labor employed as a result of the investment is the
same period after period, then there is no problem, but suppose the
time pattern of labor absorption varies from period to period, as it
may with developing skills in the use of capital; then, of course, there
is a problem similar to the one previously considered in our discussion

of the capital turnover rule. We need not go through the difficulties involved on this point, since the analysis parallels that considered under the capital turnover rule. It would follow from such an analysis, even in those cases where over the initial life of the capital good, the labor absorption rule and SMP are consistent, that they need not be consistent in the long run when all the dynamic considerations are taken into account.

Corollary (2), if taken literally, would imply that labor displacing investment would almost never be accepted no matter how high its contribution to national product.[1] For example, the United Nations experts on development tell us that:

" . . . labour-saving technology is not of great value to an economy which is over populated. There the search should be rather for technologies which increase the yield of land per acre, or which enable large numbers of persons to be employed in secondary industries for a small expenditure of capital."[2]

Investment in agricultural machinery would scarcely pass corollary (2), since it would increase the actual or disguised unemployment on the land. This is, of course, one of the implicit issues in the by now celebrated dispute between Professor Viner and his draftsman,[3] namely, that excess capacity is not an economic sin, but can be an economic virtue. Under conditions of decreasing costs the creation of excess capacity can lead to lower costs or greater output per unit of the variable factor regardless of what the fixed factor happens to be. Similarly, increasing the amount of excess labor by the introduction of labor displacing capital can result in a greater addition to output than the use of labor absorbing capital. In such a case corollaries (1) and (2) are also likely to be inconsistent with each other. The highest capital turnover rates may come about through the use of labor displacing rather than labor absorbing investment.

Finally, we may note that there is also a practical difficulty with the labor absorption rule. The extent to which the rule is useful depends on its greater ease of application than the general SMP criterion. Of what use is corollary (2), and what constitutes absorption, if the investment is to take place in an industry in which disguised unemployment exists? In order to measure the alternate points with respect to alternate investment possibilities the marginal

1. This is not to imply that any of the writers previously mentioned would suggest that the labor absorption rule be applied apart from any other considerations.

2. United Nations, *Measures for the Economic Development of Underdeveloped Countries* (New York, 1951), p. 7.

3. Jacob Viner, "Cost Curves and Supply Curves," *Zeitschrift für Nationalökonomie*, 1933.

productivity of labor in the industry (or sector) becomes zero. Surely, it would be difficult to show that this is any easier to measure than the SMP itself.[4]

It is a likely consequence of corollaries (1) and (2) that low labor productivity will be perpetuated. Wherever the SMP criterion or rules dictate the application of capital to agriculture, where there is likely to be a great deal of disguised unemployment, and where the potentialities for further population growth are usually very high, then it is highly probable that low labor productivity will be perpetuated. To begin with, the marginal productivity of labor on the land is likely to be extremely low. To the extent that investment increases productivity it will increase population and subsequently the labor force, and as a consequence *tend* to reduce average product to its former low level.[5]

The application of investment to capital-light industries such as handicrafts, corollary (1), is also likely to lead to the maintenance of low labor productivity. Since corollary (1) seeks to maximize output per unit of capital, it also implies, other things equal, using as much labor as possible with the existing capital.

We can summarize briefly why the practical corollaries that have been drawn from the SMP criterion are not appropriate for economic development. (1) The emphasis of these corollaries is on the productivity of capital not on the productivity of labor. In the long run, given the inevitable population increases that accompany industrialization, it is quite possible to raise output per unit of capital without raising output per unit of labor (labor productivity) to any extent. But it is high labor productivity that makes possible high levels of living. Put differently, corollaries (1) and (2) may yield a maximum aggre-

---

4. See Wilbert E. Moore, *The Economic Demography of Eastern and Southern Europe*, chap. III, and Appendix II, for the difficulties involved in attempts to measure "surplus" labor.

5. It may be argued that on ethical grounds, corollary (2) is sound in that it leads to lower mortality rates in agricultural areas. But this will only be a temporary gain that will be wiped out as soon as the consequent population increase reduces productivity *per capita* to its former level. Two sorts of welfare questions are involved. First, what sector should be favored with the initial increase in productivity and income? Second, is it ethical to sacrifice a temporary gain in living standards in order to insure long-run permanent gains? These are very serious questions, and welfare questions so fundamental in significance can scarcely be dealt with in a footnote. However, it may be pertinent to point out that (1) there is almost universally a direct correlation between industrialization, broadly defined, and standards of living, and (2) there is probably no case of "painless" industrialization on record, as any reader of *The Town Labourer* will recognize. Even in the United States, where an optimum situation for development prevailed, the immediate consequences were not always pleasant.

gate national product but not a maximum product *per capita.* The last two magnitudes can conceivably go in opposite directions. (2) The corollaries emphasize aggregate output and not the rate of investment, whereas it is precisely this rate that determines the extent of capital accumulation and as a consequence the capacity of the economy to produce goods and services in the future. The formal rules do not take into account what happens to the final product during any period, but it is just this that determines the rate of investment. (3) The rules enumerated above do not take into account changes in the factors other than capital. For example, no account is taken of the population growth that may be a consequence of investing in a certain manner and achieving a certain product mix. It is population and labor force growth that tend to reduce capital *per capita* and hence reduce output *per capita.* The crux of this argument is that the allocation of capital, the consequent allocation of the final product, and population growth are not independent factors.

If it be granted that the object of development is to attain a level of economic capacity which maximizes output *per capita* at a determined future time, then the correct criterion for allocating investment must be to choose for each unit of investment that alternative that will give each worker greater productive power than any other alternative. To achieve, this result we must maximize (a) the amount of capital per worker, and (b) the quality of the labor force, i.e., its skill, knowledge, energy, and adaptability.

Apart from the human factors, it is the capital/labor ratio that determines output *per capita.* From this point of view the criterion to be adopted is the one that leads ultimately to the maximum capital/labor ratio. The amount of capital per worker that is created in the long run depends on two broad factors: (1) the amount of investment year by year stemming from the product of the initial investment; and (2) the increase in the size of the labor force. We must therefore take into account the initial investment plus the sum of all subsequent reinvestments divided by the size of the labor force at the end of the stipulated time horizon.

The marginal principle applies here as it does elsewhere — but not marginal *productivity* in the usual sense. The criterion we suggest might be called the *marginal per capita reinvestment quotient.* The best allocation of investment resources is achieved by equating the marginal *per capita* reinvestment quotient of capital in its various alternative uses. The result of such a policy would be to maximize the *per capita* output potential at some future point in time.

To secure a clear notion of what is meant by the marginal *per capita* reinvestment quotient we must consider the basic factors involved in its determination. Briefly stated, the seven basic factors are as follows: (1) gross productivity per worker; (2) "wage" goods consumed per worker; (3) replacement and repair of capital; (4) increments in output as a result of noncapital using innovations, such as improvements in skills, health, energy, discipline, and malleability of the labor force; (5) declines in mortality; (6) declines in fertility; and (7) direction of reinvestment. The first six factors determine the *per capita* reinvestment available period after period, and the last one deals with its allocation.

The gross productivity per worker minus the consumption per worker determines the gross amount available for reinvestment per worker.[6] Deducting replacement and repairs per worker yields the net amount (per worker) available for reinvestment during any period. Over a succession of periods there may be increases in productivity that arise *not* out of additions to capital but rather because of increases in skill, organizational innovations, or improvement in health, energy, or discipline of the labor force that have to be taken into account. Capital per worker declines if the labor force grows at a more rapid rate than capital accumulates. But the rate of growth of the labor force will depend, for the most part, on the rates of mortality and fertility of the population, leaving aside the question of labor force participation.

It is a common experience in all underdeveloped areas that mortality declines at a faster rate and earlier than fertility as levels of consumption increase. Therefore, the greater the gap between output and consumption, the less the rate of population growth and the less the dilution of capital. The extent of fertility decline will also depend on the allocation of investment, since the type of investment that stimulates urbanization will create a more favorable environment for fertility decline than the type of investment that perpetuates the rural agricultural environment.

---

6. We abstract here from the very difficult problem of ensuring that this Ricardian "surplus" is indeed reinvested, which involves either the creation of a Schumpeterian ideology among entreprenuers, government control, or outright government investment. Which of these alternatives is adopted depends, of course, upon the specific institutional situation of the country undergoing development.

II. The Case for High Labor Productivity in Underdeveloped Areas

1. *The level of labor productivity in underdeveloped areas.*

We consider first such empirical data as are available with respect to the differences in labor productivity between developed and underdeveloped areas.

(a) An excellent statistical analysis has been made of the productivity of labor in underdeveloped areas relating to the cotton textile industry of five Latin-American countries — Brazil, Chile, Mexico, Ecuador, and Peru.[7] It was determined that for the so-called "old" mills, which comprised 90 per cent of the entire industry in these countries, "labor consumption per kilogramme of fabric (taking into account the process of spinning and weaving) is five times greater . . . than that which could be expected under the best conditions — within practical limits — of modernity of equipment, size, organization and administration."[8] The standards against which labor inputs were measured were models, to which were attributed the "best possible process and labor organization, and it was supposed that they would operate with the best efficiencies attainable without impairing the quality of the product."[9] These standards were undoubtedly above the average efficiency of cotton mills in the United States, but since they were constructed by a U. S. engineering firm, it may be assumed that good U. S. practice served as a guide. It should also be noted that cotton textile manufacturing is one of the best developed and relatively most efficient industries in Latin America.

(b) The accompanying table purports to compare productivity in

| | Output per Operative, in Pounds Sterling | | | | | |
| | United Kingdom, 1935 | | Egypt, 1937 | | United States, 1937 | |
Industry	Net value of output	Index of output	Net value of output	Index of output	Net value of output	Index of output
Chemicals.......	617	100	69	11	1,145	186
Textiles.........	159	100	39	25	318	200
Clothing.........	168	100	61	36	356	212
Leather.........	237	100	45	19	417	176
Clay and stone...	238	100	28	13	588	247
Paper...........	332	100	68	21	867	261
Food............	487	100	82	17	760	156
All manufacturing	264	100	56	21	595	225

Source: Gamal Eldin Said, "Productivity of Labour in Egyptian Industry," *L'Egypte Contemporaine*, Nos. 259–60, May–June, 1950.

7. United Nations, Department of Economic Affairs, *Labor Productivity of the Cotton Textile Industry in Five Latin-American Countries* (New York, 1951).
8. *Ibid.*, p. 5.
9. *Ibid.*, p. 127.

various Egyptian manufactures with that in the United States and Great Britain.

These figures indicate that for the years in question, productivity in Egyptian manufacturing as a whole was about 10 per cent of the U. S. level, though for the clothing industry, which is labor-intensive in the United States, the disparity was smaller.[1]

(c) Comparative productivity data for the Soviet Union in 1928, at the commencement of the industrialization program, and the United States in 1939, are shown in the following table:

SOVIET PRODUCTIVITY IN 1928 AS A
PERCENTAGE OF U. S. PRODUCTIVITY IN 1939

Coal mining	20
Iron ore mining	11
Crude oil and gas extraction	24
Iron and steel	17
Cotton cloth	16
Shoes	38
Beet sugar	14
Manufacturing and mining	22

Source: Walter Galenson, "Industrial Labor Productivity," in Abram Bergson (ed.), *Soviet Economic Growth* (Evanston, 1953), p. 190.

Soviet industry was considerably more advanced in 1928 than that of Egypt in 1937, and the difference is reflected in the comparative productivity data. Yet the Soviet 1928 level is seen to have been far below that of the United States, reflecting the great relative backwardness of the preplanning Soviet economy.

(d) A comparison of manufacturing labor productivity in China (1936) and the United States (1935) yielded the conclusion that Chinese productivity was about 5 per cent of the U. S. level.

"The difference is appalling. Yet this is not the whole picture. When the comparison is made between the output produced by an American worker and a Chinese handicraft worker, the result is almost unbelievable . . . one day's work of an American worker will be equivalent to fifty days' work of a Chinese handicraft worker. This low productivity is, of course, only partly due to the inefficiency of labor, and partly, perhaps mainly, due to the meagerness of capital investment."[2]

1. Though the data appear to have been prepared carefully on the basis of censuses of manufacturing, they are subject to the limitation of more or less arbitrary exchange rates. The author converted at the rate of 97½ piastres to the pound sterling, which prevailed in 1944, adjusted for the decline in the internal purchasing power of the piastre from 1937 to 1944. The British and U. S. data were taken from L. Rostas, "Industrial Production, Productivity and Distribution in Britain, Germany and United States," *Economic Journal*, April 1943.

2. Pao-San Ou and Foh-Shen Wang, "Industrial Production and Employment in Pre-War China," *Economic Journal*, LVI, 433.

(e) Labor productivity in Ceylon *circa* 1950 is reported to have been from 20 to 30 per cent below the level prevailing in India.[3] So far as the authors are aware, there is not available a comparison of labor productivity in India and the United States, though observers are agreed that Indian productivity is low indeed.[4]

The foregoing data are not sufficiently homogeneous to support any precise generalization as to the level of industrial labor productivity that may be expected to obtain in an underdeveloped area. For one thing, there are likely to be sharp interindustry variations. It is quite possible for single plants or industries within an industrially backward country to measure up to the levels of efficiency prevailing in advanced nations. However, the evidence that is available indicates that for manufacturing as a whole, a level of labor productivity above, say, 40 per cent of that of the United States is characteristic of a developed nation,[5] with substantially lower levels, perhaps 20 per cent and less, prevailing in underdeveloped areas. In other words, in a typically underdeveloped area, it will require at least five, and perhaps ten or even more workers to produce the same amount of goods as a single American worker.

2. *Labor productivity and capital allocation.*

From the policy interpretations of SMP that we have discussed above it would appear that India, for example, would be deviating from the prescribed course if in allocating capital to an industry, or to a process within an industry, it attempted to attain the U. S. (or British) labor productivity level unless such result were the unavoidable result of fixed technical coefficients, since the marginal product of the capital invested would undoubtedly be greater diluted than concentrated.[6] Labor productivity should presumably rise at a relatively slow rate as capital trickles into the defiles marked by marginal productivity.

In fact, actual development programs have often run counter to these prescriptions. In the case of a group of Mediterranean countries, the major share of recent investment of industrial capital

3. International Bank for Reconstruction and Development (IBRD), *The Economic Development of Ceylon* (Baltimore, 1953), p. 523.

4. See Wilbert E. Moore, *Industrialization and Labor* (Ithaca, 1951), p. 108.

5. Just before World War II, Great Britain, Germany, Russia, Sweden, and Holland were said to be within a range of 40–50 per cent of the U. S. productivity level. L. Rostas, *Comparative Productivity in British and American Industry* (Cambridge, 1948), p. 40.

6. We assume that there is a direct correlation between capital intensity and labor productivity. This proposition cannot be proved on the basis of data presently available. It has been asserted of Europe that "there is a close rela-

has gone into high capital-intensity producer services.[7] A similar allocation of capital in Mexico caused some disquiet to an apparently orthodox-minded International Bank mission.[8] A disproportionately large share of Soviet capital resources have been diverted into capital-intensive heavy industry rather than into more labor-intensive light industry, and into capital-intensive processes within industry.[9]

Some arguments in favor of the goal that we have advanced, i.e., maximization of the capital/labor ratio through application of the criterion of the marginal *per-capita* reinvestment quotient, and against the practical corollaries of SMP follow:

(a) It may be well to dispose first of the popular argument that allocation of investment to labor-intensive industry is a social desideratum where surplus population exists. For the very short run, it is incontrovertible that the maximum number of persons can be put to work with the minimum amount of capital investment if the capital is simple in form and widespread in its distribution. For example, the WPA program in the United States, the purpose of which was employment regardless of output, represented a logical application of this principle. In the longer run, however, there is considerable doubt of the general validity of the proposition.

The point may best be illustrated by a simple model. Let us assume that a product can be manufactured under alternative combinations of labor and machinery, e.g., with an automatic machine requiring little labor or with a semiautomatic machine requiring more labor. For any year, the employment provided by any combination

tionship between horse-power per head and output per head." United Nations, *Economic Bulletin for Europe*, Vol. 3, No. 1, p. 24. However, Rostas found "no correlation between horse-power per unit of output and output per worker, i.e., industries where horse-power per unit of output is higher in the United States are not identical with industries where United States output per worker is also relatively high." L. Rostas, *op. cit.*, p. 54. We should judge that the relationship between capital intensity and labor productivity is closer at lower than at higher states in industrial development, i.e., that the capital factor plays a more important role as a determinant of productivity where there is little capital than where industry is already heavily capitalized.

7. Chenery, *op. cit.*, p. 76.

8. The mission noted in its report: "In Mexico, the principal capital goods industries require a much higher investment of capital per unit of value added and per worker employed than the major consumer goods industries. It is this factor which makes the orientation of industrial investment toward capital goods industries a matter of especial concern in a country where capital resources are scarce and the potential market for capital goods smaller than for consumer-goods." IBRD, *The Economic Development of Mexico* (Baltimore, 1953), p. 65.

9. See Norman M. Kaplan, "Capital Formation and Allocation," in Abram Bergson, *op. cit.*, p. 37, and the comments of Alexander Erlich, *ibid.*, pp. 92–97.

of men and machines, can be represented by the following equation:[1]

$$(1) \qquad E_{t+1} = E_1 \left(1 + \frac{p - ew}{c}\right)^t$$

1. The variables may be defined as follows:

I = total investment in any period
P = gross value added in any period
W = total real compensation of labor in any period
w = real wage rate
N = number of machines
p = output per machine, i.e., $P = Np$
E = total employment
e = number of workers per machine, i.e., $E = eN$
c = cost per machine
v = wage cost per machine, i.e., $v = ew$

It is assumed that $I = P - W$, i.e., that the total amount invested in any period is the difference between total gross value added and the real compensation of labor. Then:

(1.1) $I = P - Ew$

(1.2) $I = Np - Ew$

(1.3) $I = Np - eNw = N(p - ew)$

(1.4) $c\Delta N = N(p - ew)$

$$\Delta N = \frac{N(p - ew)}{c}$$

(1.5) $\dfrac{\Delta E}{e} = \dfrac{N(p - ew)}{c}$, from the relationship $\Delta E = e\Delta N$

(1.6) $\Delta E = \dfrac{eN(p - ew)}{c}$

(1.7) $\Delta E = \dfrac{E(p - ew)}{c}$, and substituting $v = ew$,

(1.8) $\Delta E = \dfrac{E(p - v)}{c}$

This yields the basic equation: $\Delta E_t = E_t \left(\dfrac{p - v}{c}\right)$

(2.1) $E_{t+1} = E_t + \Delta E_t = E_t \left(1 + \dfrac{p - v}{c}\right)$

(2.2) $\dfrac{E_{t+1}}{E_t} = 1 + \dfrac{p - v}{c}$ for all t

(2.3) $E_t = E_1 \dfrac{E_2\,E_3}{E_1\,E_2} \cdots \dfrac{E_{t-2}\,E_{t-1}\,E_t}{E_{t-3}\,E_{t-2}\,E_{t-1}}$

(2.4) $E_t = E_1 \left(1 + \dfrac{p - v}{c}\right)^{t-1}$

This may be rewritten

(2.5) $E_{t+1} = E_1 \left(1 + \dfrac{p - v}{c}\right)^t$, and in our original symbols

(2.6) $E_{t+1} = E_1 \left(1 + \dfrac{p - ew}{c}\right)^t$

where E_{t+1} represents employment in the $t+1$ year, E_1 employment in the initial year, p the output per machine, e the number of workers per machine, w the wage rate, and c the cost per machine. If E_1, p, and e are assumed to be parameters, then the value of E_{t+1} will depend upon the relationship of c and w.

Essentially, all this formula does is to state formally a simple notion that can be grasped intuitively: that the larger the portion of the output of an industry, or a society, which is reinvested rather than consumed, the more rapid will be the process of capital accumulation, and *pari passu*, the growth of employment opportunities in industry. Conversely, the greater the share of output that is consumed, the slower will be the rate of expansion of capital and employment.

The following data for the Indian textile industry *circa* 1943, may be used to illustrate the application of the formula:[2]

	Capital investment per worker (rupees)	Value added per worker (rupees)	Annual earnings per worker (rupees)
1. Power machinery, large scale .	1200	650	80
2. Power machinery, small scale .	300	200	80
3. Automatic loom, cottage industry..........	90	80	80
4. Hand loom, cottage industry .	35	45	80

So, for example, if we apply the data for a small scale power loom to formula (1), we have, for year $t+1 = 5$, the following:[3]

$$E_5 = 4\left(1 + \frac{200 - (1)\,(80)}{300}\right)^4 = 15.3$$

Similarly, the formula may be applied to the above data for other types of mills, and for other time periods, the results being as follows:

2. The source of the capital and value added data is *The Eastern Economist*, July 23, 1943, p. 340. The figure of 80 rupees used in the illustration was purely hypothetical, chosen for purposes of illustration. In fact, annual earnings in the non-cottage sector of the Indian textile industry were considerably higher in 1943, varying from 204 rupees in Bihar to 832 rupees in Bombay. A. N. Agarwala, *Indian Labor Problems* (Allahabad, 1947), pp. 49–50. However, the capital data in the example may well refer to an earlier year, and since there was a severe wartime inflation, a realistic wage figure is difficult to determine. Moreover, the wage should be differentiated among the various sectors of the industry.

3. In applying this formula, it is assumed that total investment in each case is 1200 rupees, so that in the case of the small scale power loom, an initial investment of 1200 rupees would make $E_1 = 4$. In each case e is taken to equal one, and therefore p is equal to the value added per worker.

HYPOTHETICAL EMPLOYMENT PROVIDED BY INITIAL
INVESTMENT OF 1200 RUPEES IN VARIOUS TYPES OF COTTON
TEXTILE MACHINERY

Year (t + 1)	Modern mill, large scale	Power loom, small scale	Automatic loom, cottage industry	Hand loom, cottage industry
5	5	15	13	35
10	34	83	13	35
15	242	444	13	35
20	1,718	2,390	13	35
25	12,200	12,860	13	35

It is assumed in the above example that the total product, less labor cost, is reinvested each year. No allowance is made for capital replacement, although this could be accomplished by reducing the product available for reinvestment by a depreciation factor, thus slowing up the process of capital accumulation. As we shall see, however, it may be quite appropriate to disregard the depreciation factor in an economic development model of this character.

It is clear from the above example that the wage rate (w) is the critical variable; that the amount of employment provided in any future year by a unit of investment depends largely upon the wage deduction from the product of industry. If, for example, an annual wage of 150 rather than 80 rupees had been used,[4] employment provided by the large scale modern mill would have been 24 in ten years, as against only 16 in the small scale power mills. The higher the real wage level, the greater is the advantage in terms of potential employment accruing from the use of capital-intensive machinery.[5]

(b) Failure to introduce capital intensive techniques at the outset of the industrialization process may create insurmountable institutional barriers to modernization. This is true particularly in democratic communities, where labor/capital coefficients cannot be altered by fiat. The case of Cuba is very much in point:

"Cuban industrial development has been greatly retarded by labor's resistance to new machinery, modern methods, or virtually anything that will increase the efficiency of production. . . . When improved methods or machinery are introduced into a factory with permission of the workers, it is generally under the stipulation that the same number of workers be employed as were used under the older, inefficient methods. The workers also commonly see to it that the new equipment turns out no more products than the old."[6]

4. See *supra*, n. 2, p. 358.
5. Thus in Surinam, where wages are relatively high, it was suggested that "the development of the country must rather be in the direction of relatively high capitalized production, where the productivity of the workers can be great enough to justify their level of wages." IBRD, *Surinam* (Baltimore, 1952), p. 26.
6. IBRD, *Report on Cuba* (Washington, 1951), pp. 143–44.

In neighboring Puerto Rico, "the legislative branch of the Insular government, which is highly responsive to the representations of the electorate, has shown considerable interest in creating and protecting job opportunities, although these sometimes involve make-work practices for a class of firms which might otherwise be unnecessary."[7]   Low productivity in Mexican textile manufacturing was attributed in large measure to "present contract legislation for the industry, which stipulates the number of workers to be employed in relation to the capacity of the mills, and establishes an inflexible basis for the proportion between production and wages."[8]

In all of these countries it was observed that newly established factories had considerable latitude in determining their labor/capital ratios; it was in the older, established plants that the problem was acute.   This suggests that the failure to adopt the correct criterion at the very outset of a development program may render infinitely more difficult the attainment of high output per worker, which must be the eventual goal of a development program.

(c) Industrialization almost inevitably means urbanization, and urbanization entails charges upon the production surplus available for capital accumulation.   In some countries, e.g., Turkey, an industrial labor force must be recruited from the farm and brought to the city, entailing a considerable investment in such urban facilities as housing, sanitation, water and canteens.   But even where large cities with surplus manpower already exist, as in India or China, the social costs involved in bringing newly recruited industrial workers up to a minimum efficiency level may be very considerable.   The Colombo Plan, for example, allocated 18 per cent of projected capital investment to "social capital," compared with only 10 per cent to "industry and mining."[9]   In the case of Egypt, it has been pointed out that many workers are living below the subsistence level, in completely inadequate houses, leading to chronic undernourishment, and high disease rates, which in turn result in a high rate of absenteeism and low productivity.   Where an efficient labor force is to be maintained, the social capital costs may be very high.   Such costs are often neglected when the criterion of SMP, and particularly its accompanying corollaries, are advocated.

The establishment of the highest initial productivity of labor will

7. Simon Rottenberg, "Labor Cost in the Puerto Rican Economy," *Revista Juridica of the University of Puerto Rico*, Vol. XX, No. 2, Nov.-Dec. 1950, p. 59.
8. *Labor Productivity of the Cotton Textile Industry in Five Latin-American Countries*, op. cit., p. 84.
9. Chiang Hsieh, "Underemployment in Asia," *International Labor Review*, July 1952, p. 32.

minimize urbanization cost by bringing into the industrial labor force a minimum of workers. One may question the wisdom of *minimizing* the transfer of manpower from agricultural employment to industrial and urban pursuits, but the optimal transfer rate would probably be significantly lower than the maximum rate of labor absorption. The establishment of a high initial productivity of labor, following our criterion for allocating capital, will yield a lower ratio of urbanization cost to output per worker than under the orthodox criterion, and thus contribute to maximizing the rate of capital formation.

(d) Too little attention has been paid to the *pattern* of industry which will facilitate, or indeed, make possible, industrial development. It is not a matter of indifference whether capital is allocated, say, to the manufacture of iron and steel, or to the manufacture of textiles. Development mission recommendations are all too prone to assume that funds invested in light industries based, for example, on local raw materials, will in some unexplained manner lead eventually to economic development.

Properly speaking, neither the criterion of SMP in its orthodox interpretation nor our SMP criterion will lead automatically to the establishment of an industrial pattern which will maximize industrial development. This is a subject as yet largely unexplored, and it can probably best be tackled in terms of a model investment grid, based upon the specific conditions of each underdeveloped area. However, the above-cited corollaries of SMP are likely to lead in a direction other than industrialization, for the average productivity of capital tends to be greatest in light industries, where a considerable amount of labor may be used, and least in heavy industries, where the substitutability of labor for capital is less.[1] Conversely, our criterion of maximizing the capital/labor ratio tends to favor those industries which are essential to the development of modern industry.

(e) As we have pointed out, the usual prescriptions for capital allocation favor short-lived over long-lived capital goods. From the point of view of our suggested criterion a strong case can be made for the obverse rule; viz., that there are considerable advantages to be

1. This is an empirical statement for which factual proof is largely lacking. However, 1947 U. S. Census of Manufactures data indicated that capital investment per worker in American industry clearly tended to be greater in the heavy capital goods industries than in the light consumer goods industries. For example, capital investment per worker in the industrial chemicals industry was $15,868; in iron and steel, $7,309; in machinery (excluding electrical), $6,993; compared with $4,673 per worker in textile mill products; $2,954 in apparel; $3,376 in leather and its products. There were exceptions to the rule, such as canning and preserving, with $10,036 invested per worker, but in general the data appear to support the statement in the text.

gained from longer-lived rather than shorter-lived capital goods. Domar has demonstrated that under conditions of growth, replacement/gross investment ratios can be considerably lower for longer-lived than for shorter-lived capital goods.[2]

At the outset it should be made clear that it is replacement and not depreciation that is the significant variable for purposes of development. Depreciation cost, as usually conceived, is an accounting matter,[3] and is not directly connected with the productive power of a capital good at any point in time. Thus, for our purposes the important variable is not net investment, as usually defined, but gross investment less replacement cost. (For ease in communication we shall refer to gross investment less replacement as GIR.)

Initially there are three advantages to longer-lived capital under conditions of capital growth. (1) The longer the life of the capital, the longer the period of time during which no replacements have to be made, and hence the greater the available output per man during this period than with capital of a shorter life. (2) Out of a greater output per man there is the possibility of greater reinvestment per man. (3) With greater output and reinvestment per man there is a greater chance of overcoming the critical minimum effort[4] hurdle than otherwise.

If the capital stock is to be maintained intact, then eventually replacements must be made. But even after the advantage of the early no-replacement period has passed, there is still an advantage in a capital stock that is on the average longer-lived than shorter-lived under conditions of growth. This is because the longer the average life of the capital, the smaller is the proportion of gross investment needed for replacement, and therefor the larger is the GIR period after period. For example, suppose that the rate of gross investment growth is 5 per cent per year, then, according to Domar's figures,[5] if the average length of life of capital is four years, the ratio of replacement to gross investment will be 82 per cent per year; if the average length of life of capital is ten years the replacement/gross investment ratio drops to 61 per cent; and if the average length of life of capital is thirty years, the replacement/gross investment ratio drops as low as 22 per cent. Clearly, the greater the life span of

2. E. D. Domar, "Depreciation Replacement and Growth," *Economic Journal*, LXIII (March 1953), 1–32.

3. Strictly speaking, physical depreciation interpreted as user cost is an economic variable that does have to be accounted for. In order not to complicate the argument we assume that user cost is an insignificant portion of accounting depreciation. In any event, user cost is a short-run cost consideration since it varies with output.

4. See below, p. 366, for a discussion of this term.

5. Domar, *op. cit.*, p. 8.

investment goods, other things equal, the greater the possibility for economic growth.

3. *Population growth and development.*

(a) To a considerable extent our arguments in the previous sections were based on the notion that there is a dependent relationship between the allocation of investment, the consequent allocation of the labor force, and the growth and quality of the population. Although this is hardly a novel idea it is customary for economists to argue as though these were independent matters.[6] Solutions that may appear to be appropriate in instances where the population problem is disregarded often cease to be appropriate when it is recognized that the population aspect is an integral and inseparable part of the problem. Although we cannot go into detail at this juncture, we shall indicate something of the nature and significance of this point.

In considering the population factor there are three things about which we can be fairly certain on the basis of past experience: (1) that economic development, if it takes place, will be accompanied by rapid population growth; (2) that the only way to reduce population growth is to reduce the birth rate; and (3) that declines in birth rates, if they take place, will follow rather than precede declines in death rates. This implies that in order to avoid the Malthusian dilemma, and at some point to achieve a slackening of population growth, it is necessary to create an environment that is conducive to a reduction in birth rates. From past evidence it would appear that the urban, industrially and commercially developing sectors rather than the rural agricultural areas provide the environment conducive to falling birth rates.

Reference to the vital statistics of most underdeveloped areas suggests that economic development will be accompanied by rapid population growth. On the average the underdeveloped countries of the world are increasing their population at a rate of 1.4 per cent per year.[7] Improved nutrition and public health measures that often accompany economic development can reduce death rates so that the average rate of growth may rise to 2 or 2.5 per cent per year. The last rate will yield an increase in population for the underdeveloped

---

6. For example, Professor Viner, in observing that population increases will tend to worsen a country's terms of trade, argues that " . . . this will apply equally, as a tendency, to countries whether they are predominantly agricultural or predominantly industrial, and the appropriate remedy in either case would be to check the rate of population growth." (Viner, *op. cit.*, p. 142.)

7. This generalization is based upon a study of the United Nations, *Demographic Yearbook*, 1953.

two-thirds of the world of some two *billion* people by 1984. So much for the magnitude of the problem.

The reasons behind the other assertions have been investigated and long studied by demographers and ably summarized by Professor Notestein:

"The more rapid response of mortality than of fertility to the forces of modernization is probably inevitable. The reduction of mortality is a universally acceptable goal and faces no substantial social obstacles. But the reduction of fertility requires a shift in social goals from those directed toward the survival of the group to those directed toward the welfare and development of the individual. This change, both of goals and of the social equipment by which they are achieved, is at best a slow process. As a result, the period of modernization is virtually certain to yield rapid population increase."[8]

On the factors that were and are responsible for the eventual reduction in birth rates Notestein writes that they

" . . . center around the growing individualism and rising levels of popular aspiration developed in urban industrial living. With the growth of huge and mobile city populations, the individual came to depend less and less on the status of his family for his place among his fellows. The station to which he was born gave place to his accomplishments and possessions as the measure of his importance. Meanwhile, the family lost many of its functions to the factory, the school, and commercial enterprises. All these developments made large families a progressively difficult and expensive undertaking; expensive and difficult for a population increasingly freed from older taboos and increasingly willing to solve its problems rather than to accept them. In short, under the impact of urban life, the social aim of perpetuating the family gave way progressively to that of promoting the health, education, and material welfare of the individual child; family limitation became widespread; and the end of the period of growth came in sight."[9]

If Notestein's view is correct, and it is certainly true for the past, then the allocation of investment funds can be important in determining the point at which the rate of population growth declines. Certainly, investment in agricultural-rural pursuits will have much less effect (if not a negative effect) in fostering fertility declines than the creation of an urban industrial-commercial environment where the factors mentioned by Notestein can take root, grow, spread, and have their effect on the birth rate.

At first blush there may appear to be an apparent contradiction between the point just made and our previous recognition of the problem of the social cost of transferring workers from agriculture to urban pursuits. We cannot suggest that on the one hand, city-ward migration should be reduced in order to minimize the social costs of

8. Frank W. Notestein, "Population — The Long View," in Theodore W. Schultz (ed.), *Food for the World*, pp. 40–41.

9. *Ibid.*, p. 41.

urbanization, and at the same time that rural-urban migration be maximized in order to facilitate the reduction of fertility rates. But this is only an apparent dilemma. Both the fertility effect and the social transfer costs effect must be taken into account and balanced against one another in terms of their combined effect on the eventual capital/labor ratio. These two factors are not entirely antithetical, for it is rising standards of living as a consequence of increasing wages and productivity in the urban environment that creates the proper atmosphere for fertility decline. If wages in the new industries are permitted to decline through the substitution of labor for capital, this will not only reduce the rate of reinvestment but also diminish the creation of those social conditions that lead to the economic and social mobility conducive to fertility decline. Thus, an optimum rate of urbanization should not be confused with the short-run maximum rate; it is rather the highest rate consistent with the maintenance of increasing wages and productivity under general social conditions conducive to fertility decline.

There is another viewpoint from which the two factors are not contradictory. The net social transfer cost is the opportunity foregone of adding to the stock of producers goods. But that part of the social transfer cost that aids in fertility reduction makes an indirect contribution toward increasing the capital/labor ratio. Thus, expenditures on educational facilities, birth control clinics, or other facilities that change attitudes or dispense information conducive to fertility decline contribute toward increasing capital per worker in the future. Much more can be said on this aspect of our problem but we would certainly go beyond the bounds of this paper if we attempted to derive a theory of the optimal rate of urbanization.

Our interest in the possibility of achieving varying rates of population growth as a result of different investment patterns arises out of the consequences of varying rates of population growth on the *per capita* output potential. First, population growth reduces the capital/labor ratio and as a consequence reduces the *per capita* output potential. Second, population growth, through its adverse effect on *per capita* output and savings, tends to reduce the rate of reinvestment as a consequence of any initial investment. Thus, the lower the rate of population growth, or the earlier the decline in the rate of population growth sets in, the greater the *per capita* reinvestment quotient. In sum, the allocation of investment will affect not only total output but also the distribution of the labor force, the growth of the urban sector of the community, the social and cultural conditions under which people live, and the consequent attitudes towards early mar-

riage, family size, and the resultant population growth. Therefore, the allocation that maximizes current output may be quite different from the allocation that maximizes the ultimate capital/labor ratio or the ultimate output potential per man.

(b) If an underdeveloped country is to develop successfully, it is necessary for that country to make a large initial effort to increase output and to do so very early in the development attempt. If the initial or early effort does not reach a critical minimum, then it is likely that the country will revert back to its former underdeveloped stage. The reasons for the need of an initial or early minimum effort follow:

(1) The potential growth of population makes it necessary to have initially a sufficiently large increase in capital, which induces successive increases in capital, so that it is not possible for the population increases to reduce the average amount of capital per worker.

(2) A minimum effort is necessary in order to create the economic environment where external economies are possible. For example, railroads, communication systems, and irrigation works all may require large-scale initial efforts.

(3) The initial effort must be sufficient to raise incomes per head so that savings can be achieved, to continue the rate of capital accumulation. For once the rate of capital accumulation declines, there is always the challenge of population increases reducing the amount of capital per head.[1]

The critical minimum investment necessary will depend, in part, on the allocation of the investment funds. The SMP corollary allocation may or may not minimize the critical minimum investment necessary, while the allocation under our criterion, which directly takes into account the population factor, will tend to maximize the chances of overcoming the critical minimum. For that allocation that yields a lower rate of population growth, or leads to an earlier decline in the rate of population growth, reduces the critical minimum investment necessary, other things equal.

We may note that the critical minimum effort thesis implies the desirability of rapid capital accumulation early in the process of development not only in order to overcome the population hurdle,

---

1. For an elaboration of this idea see H. Leibenstein, *A Theory of Economic-Demographic Development* (Princeton University Press, 1954), chaps. IV and V. H. W. Singer appears to suggest the same idea, in somewhat different words, when he argues that the underdeveloped economics are faced with a series of interlocking vicious circles that only an able, significant, and sustained effort can break. "Economic Progress in Underdeveloped Countries," *Social Research* (Mar. 1949), pp. 5 ff.

but also because there may be a connection between the tempo of change, the rate of urbanization, and the creation of an environment conducive to the lowering of the birth rate. This last point, while not an established fact, would appear to be a reasonable speculation since growing individualism, economic and social mobility, loosening of family function, and lesser dependence on family status and traditional values generally go hand in hand with a quickening economic tempo.

4. *Some obstacles to high labor productivity in backward areas.*

The determination of a correct general criterion for capital allocation is not tantamount to its application. It may well be that environmental and institutional factors dictate extensive modifications in practice.[2] For example, even if it were determined that an indigenous iron and steel industry was basic to rapid development, the lack of iron ore and fuel might prevent its establishment. There may be a hundred reasons why a high initial level of labor productivity is impossible of achievement.[3]

It is our belief that the criterion of the marginal per capita reinvestment quotient and the corollary of a high capital/labor ratio are appropriate general guides in programs for economic development. In this section we shall consider some of the factors which stand in the way of the indicated solution for allocating available investment resources.

(a) The surplus of manpower, and the consequently low wage level, that characterize the major underdeveloped areas of the world, are serious stumbling blocks in the path of economic growth. The surplus may be hidden in the form of underemployment in agriculture, where the marginal productivity of labor must often approach zero; or it may be all too evident in the form of chronic urban unemployment. In extreme cases, low quality labor may be virtually a free good; in Ecuador, for example, it was estimated that a cotton mill could have ten times as many workers as an American mill without any difference in the relation of the labor cost to the value of the finished product.[4] "Because of keen competition in the employment market, the levels of wages earned in these occupations are kept extremely low, and because of low wages the management has little

2. Also moral or religious considerations may require modifications of the criterion.

3. See Wilbert E. Moore, *op. cit.*, chap. V.

4. *Labor Productivity in the Cotton Textile Industry in Five Latin-American Countries, op. cit.*, p. 67.

incentive to raise its standards of efficiency."[5]   In such situations the employment of as large a number of workers as possible appears not only to make good economic common sense, but to be socially desirable as well.

One possible solution under these circumstances is to alter conditions to conform with our criterion by making labor scarce artificially. This can be done in a number of ways: by legislation establishing relatively high minimum wages and working conditions;[6] by direct governmental control of manpower; or, in the case of state industry, by imposing high labor productivity targets upon management. None of these prescriptions is an easy one to follow.   The islands of favored employment will have to be protected by the government in some manner, for individual entrepreneurs will find it difficult to resist the constant temptation of cheap labor.   Yet there is ample precedent for such a policy; one may cite the oil refineries in the Near East, where foreign management polices the barriers, and a modern Egyptian textile mill where native entrepreneurial policy is effective in this regard.[7]

(b) The shortage of skilled labor is almost invariably cited as an obstacle to high productivity,[8] and undoubtedly it is.   Yet there is reason to believe that the difficulties in this respect are exaggerated. Looking to the Russian industrialization experience, it was necessary to train for the semiskilled trades millions of raw farm hands within the space of a few years.   During the First Five Year Plan the Soviet industrial labor force roughly doubled in size.   Of the 12.5 million new workers who entered industry for the first time between 1928 and 1932, 8.5 million had been peasants.[9]   The training they received was certainly not thorough; most of it was done directly on the job.   True

5. Chiang Hsieh, "Underemployment in Asia," *International Labor Review*, June 1952, p. 703.

6. It is quite true that higher wages by themselves reduce the amount of production available for reinvestment.   The point is that higher wages, by restricting the use of manpower, may lead to a lower *total* wage bill, depending upon the elasticity of the demand for labor and the absolute magnitudes involved. Furthermore, by developing a socio-economic setting in which we can overcome the institutional barriers to rising productivity we can get a larger surplus for reinvestment.

7. Gamal Eldin Said, *op. cit.*, p. 506.

8. E.g.: "Even more serious, there will in all likelihood be severe shortages of skilled labor for jobs which require technical knowledge, or which involve the exercise of organizational or supervisory talent."   IBRD, *Surinam*, p. 89. "Other difficulties may well emerge in the supply of foremen, skilled and semi-skilled workers."   IBRD, *The Economic Development of Iraq*, p. 83.   " . . . the acute shortage of trained personnel . . . is one of the main obstacles to rapid progress." IBRD, *The Economic Development of Ceylon*, p. 55.

9. Solomon Schwarz, *Labor in the Soviet Union* (New York, 1951), pp. 9, 31–32.

vocational schools were virtually nonexistent. Yet though the process was wasteful, it sufficed, and permitted the Russians to take greater pains with subsequent generations of labor inductees. The training of technical personnel was a longer-range problem, and here the cadres of well-trained pre-Soviet engineers were of great importance. By 1940, however, technical schools were turning out large numbers of qualified engineers and technicians.

(c) Under contemporary conditions of economic development in the democratic countries, a strong labor movement is likely to arise at the inception of industrialization, rather than at a fairly advanced stage as in the past. Moreover, the unionism of backward areas is apt to be radical in its orientation, in view of poor labor conditions and the impatience of workers with the seeming inability of the method of collective bargaining to secure for them immediate betterment. There is therefore ever present the danger that resources will be diverted from investment to consumption, that the seeds of industrial development will not be permitted to mature.

High productivity, with its implication of fewer and better paid workers, will tend to mitigate the upsurge of labor protest, to moderate the extremism of the labor movement, and to provide a climate of worker opinion favorable to technological change.[1] This is by no means a complete answer to what will undoubtedly be one of the crucial problems in industrialization, but at least it can be said that the high productivity solution appears to provide the labor conditions which are most propitious for success.

III. Conclusion

We have endeavored to show in the foregoing pages that the criterion of allocating investment on the basis of the marginal productivity of each unit of capital invested is not suitable for contemporary underdeveloped areas because of the invalidity of what we have termed the *ceteris paribus* assumption. When the facts of rapid population growth, political instability, and institutional obstacles to technological change, all of which are generally typical of underdeveloped areas today, are taken into account, it becomes clear that time is of the essence in developmental programs. The process of development must be sufficiently rapid to satisfy the swiftly burgeoning aspirations of people suddenly released from a Malthusian world and endowed with political power. The alterna-

---

1. We assume that the low productivity solution will yield such small results in terms of higher standards for the larger number of workers affected as to have little mollifying effect upon the degree of labor protest.

tive is the constant encroachment of consumption upon the national product; falling rather than rising *per capita* incomes as a consequence of explosive population growth; and eventual chaos as the mass organizations called into being by industrialization in "width," rather than in "depth," are forced to exploit their power by pressure from below.

Our thesis, baldly, is that successful economic development under present conditions, particularly in the face of gross backwardness, hinges largely upon the introduction of modern technology upon as large a scale as possible. Professor Gerschenkron has pointed out that in the past,

> "to the extent that industrialization took place, it was largely by application of the most modern and efficient techniques that backward countries could hope to achieve success, particularly if their industrialization proceeded in the face of competition from the advanced country. . . . This seems to explain the tendency on the part of backward countries to concentrate at a relatively early point in their industrialization on promotion of such branches of industrial activities in which recent technological progress has been particularly rapid. . . . In viewing the economic history of Europe in the nineteenth century, the impression is very strong that only when industrial development could commence on a large scale did the tension between the pre-industrialization conditions and the benefit that may be expected from industrialization became sufficiently strong to overcome the existing obstacles and to liberate the forces that made for industrial progress."[2]

This applies with even greater force to twentieth century Asia and Africa, with their far greater relative and absolute backwardness. These areas can best hope to see the completion of successful programs of economic development within the reasonable future, and (what is critical) under the auspices of political democracy, if in allocating available capital resources, the twin desiderata of up-to-date equipment and relatively high initial capital/labor ratios are kept to the fore.

WALTER GALENSON.
UNIVERSITY OF CALIFORNIA   HARVEY LEIBENSTEIN.
BERKELEY

2. Alexander Gerschenkron, "Economic Backwardness in Historical Perspective," in Bert F. Hoselitz (ed.), *The Progress of Underdeveloped Areas* (Chicago, 1952), pp. 7–9.

DIFFERENTIAL RATES OF PRODUCTIVITY GROWTH AND INTERNATIONAL IMBALANCE*

By J. M. Letiche

I

If only because of its sheer economic weight, the economy of the United States plays a key role in the mid-twentieth century trading world. Though the widespread use of the term "dollar shortage" is of fairly recent origin, the persistent export balance of the United States of which the "dollar shortage" is the monetary manifestation is clearly a long-term phenomenon. The mechanism of adjustment *qua* mechanism is monetary but the fundamental variables which it is expected to adjust are *real* variables that may be unresponsive to the monetary pressures of the mechanism over long periods of time. This unresponsiveness may be the result of dynamic forces, such as long-term economic development, which maintain persistent pressures opposite to those of the adjusting mechanism.

In order to determine the nature of these persistent pressures, we shall present and evaluate a series of hypotheses designed to explain long-term international imbalance. By testing these hypotheses with domestic and international trade data, we shall attempt to determine which of these forces apply generally and which apply only to individual cases, which forces are structural, resulting from growth and change, and which are institutional, resulting from abnormal conditions and ill-advised policies. The conclusions which follow from this analysis help to resolve the relative importance of the "structural" and "institutional" factors currently operating in a balancing or imbalancing direction, and to suggest the consequences of alternative policies in dealing with them.

II

It is generally maintained that over the past few decades technological progress in the United States has exceeded that in other parts of the Western world, with the result that she has led in increasing productivity.[1] Even though no other obstacles had stood in the way of the effective functioning of the mechanism of adjustment, this continuing relative reduction in real costs of production within the United States would have offset the inflationary price and income effects of the export balance which is an essential part of the adjusting mechanism. This offsetting influence, so the argument runs, was the more marked because the very industries in which the United States was making the greatest technological improvements were the industries which were highly competitive with many important export commodities of leading industrial countries abroad, export commodities which were greatly dependent upon the American market. Had technological improvement in the United States proceeded at a more even rate between industries, and had a similarly even rate of progress occurred abroad, either deflation of prices and wages abroad or a devaluation of some foreign currencies as called for by the theory of the mechanism of adjustment would have been more effective in preventing the balance-of-trade disequilibrium. In other words, had rates of technological improvement been more uniform as between industries in the United States and abroad, appropriate degrees of devaluation or deflation sufficient to keep up with the greater rate of technological progress in the United States might have worked along the lines of comparative advantage.

III

Professor Hicks has presented a very useful analytical frame of reference for dealing with this problem.[2] The substance of his position may be summarized as follows. Consider two countries, *A* and

---

1. This phenomenon of differential rates of productivity growth has been stressed by many economists as a fundamental cause of long-term international imbalance of western Europe vis-à-vis the United States. Cf. John H. Williams, *Economic Stability in the Modern World* (University of London: Athlone Press, 1952), especially pp. 9–15; J. R. Hicks, "An Inaugural Lecture," *Oxford Economic Papers*, n.s., V (June 1953), 121–35. For more general treatments cf. Charles P. Kindleberger, *The Dollar Shortage* (New York: John Wiley and Sons, Inc., and Technology Press of M.I.T., 1950), chaps. 7 and 8, and T. Balogh, *The Dollar Crisis: Causes and Cure* (Oxford: Basil Blackwell, 1950).

2. Hicks, *op. cit.*

B, which are trading together and which are such that productivity in *A* is increasing uniformly more rapidly than productivity in *B*. Hicks assumes initially that productivity in *A* is increasing uniformly among industries and steadily through time while productivity in *B* remains constant. The problem concerns the nature of long-term adjustments to such an enduring disturbance.

Hicks begins from a position in which payments between the two countries are in balance. He assumes further that money incomes in *B* remain the same, while money incomes in *A* rise to the full extent of the rise in productivity. Then the real effects of the development are most unlikely to be harmful to *B*. The *uniform* increase in productivity in *A* would tend to lower costs and prices of *A*-products, but this effect would be wiped out by the rise in *A*-incomes, so that the prices of *A*-products would be unchanged. The prices of *B*-products would also be unchanged. Consequently *B* would buy neither more nor less from *A* than before. But *A*-income has risen, and with all prices the same, this would probably cause *A*'s demand for *B*-exports and *A*'s demand for *A*-exports to rise. The balance of payments would consequently turn in *B*'s favor. This, however, would be inconsistent with long-term balance. Hicks therefore states that if the balance of payments is not to turn in *B*'s favor but to remain in balance (as it must in the accounting sense), the rise in *A*-incomes must be less than the rise in productivity.[3] If *A*-incomes rise by less than the rise in productivity, there will be some reduction in the prices of *A*-exports. Thus money income in *B* can remain constant, and the prices of *B*-products can remain constant; nevertheless the balance of payments can remain balanced, while *B* gets her imports from *A* more cheaply than before. The real income of the *B*-population must therefore be improved. So far as the real effects are concerned, a uniform increase in productivity in *A* is almost certain to affect *B* advantageously.

However, the monetary effects may be entirely different. Assuming a stable rate of exchange between *A*'s and *B*'s currency, if *A*'s productivity rises faster than *B*'s, and *A*'s money incomes do not rise correspondingly faster or nearly so, balance-of-payments crises will arise in *B* because its prices will get out of line. There will, however, be some appropriate rate of increase in *A*-incomes which will keep trade between the two countries in balance, even though the level of *B*-incomes remains in money terms completely unchanged. If *A*-incomes do rise at this appropriate rate, then there need be

3. The validity and relevance of this proposition will be discussed below.

no deflation in B, while B is benefited by better commodity terms of trade.

Hence, when the rate of growth in productivity is more or less *uniformly* distributed among all industries, and is higher in one country than in another, then the country with the less rapid rate of growth may encounter monetary difficulties; but *if* these can be overcome, some part of the gain of the more progressive country will go to the less progressive one and no mechanism-of-adjustment difficulties need arise.

IV

Hicks draws particular attention to the fact that if the rate of growth of productivity is not uniformly distributed, structural as well as monetary difficulties are likely to arise. Improvements in productivity may be what he calls "export-biased" or "import-biased." Assuming that money incomes remain the same in both countries, if A's productivity in export industries is rising relatively rapidly and its productivity in import-competing industries is rising relatively slowly, the double factoral terms of trade will remain unchanged and commodity terms will improve for B. When the improvement is concentrated upon goods which A exports, and when A-money incomes and B-money incomes remain the same, the general level of factor prices must remain the same; consequently, country B shares fully in the improvement through the lower export prices of A. These circumstances are less likely to lead to monetary difficulties in B; they are less favorable to the progressive, improving country A, and more favorable to B.

But, according to Hicks, if A's productivity in export industries is rising relatively slowly and its productivity in import-competing industries is rising relatively rapidly, not only will the double factoral terms of trade turn against B (creating the same *monetary* difficulties as in the case of uniformity), but the commodity terms will also turn against B, and this will tend to bring about a lower level of its real income. Consider the case of a uniform improvement in those A-industries which compete most closely with B-exports, but no improvement in other A-industries, and none at all in country B. Then, if money incomes remain the same in both countries, the prices of A-products which compete with B-exports will fall. A's demand for B-exports will decline. This development would be enough to cause a deficit in B's balance of payments, for there is no reason, says Hicks, why B's demand for A-exports should be affected. In order

to maintain a trade balance there must be a rise in *A*-incomes relatively to *B*-incomes, i.e., the double factoral terms of trade must move adversely to *B*. Further, since there has been no improvement in productivity in *A*'s export industries, the rise in *A*-incomes implies a rise in the prices of *A*-exports. This tendency, plus the competitive pressure upon *B*-exporters to lower their prices, must turn the commodity terms of trade against *B*.[4] Thus, concludes Hicks, an improvement in *A*-productivity that is concentrated upon those *A*-industries which compete most closely with *B*-exports must make *B* worse off — regardless of the course of money incomes and irrespective of the type of international monetary arrangements.

For the anlysis of historical and contemporary experience with respect to international imbalance, a determination of the type of improvements made by the respective countries is therefore a primary requisite. Hicks states that the case of *uniform* improvements is not applicable to the problem of long-term external imbalance of western European countries vis-à-vis the United States. His explanation of the long-term "dollar-shortage" problem runs as follows:

During the nineteenth century, differential rates of growth in productivity were mainly concentrated in the export fields on both sides of the Atlantic, though the average rate of such progress probably varied among different countries. The general characteristic of British-American trade, for example, was an exchange of the products of American farms for those of British factories. The high productivity of American manufacturing began very early. But British manufacturers were able to compete with Americans on the basis of lower wages. Similarly, the high wages in the United States did not prevent Britain from importing American farm products advantageously to herself, because the exceedingly high productivity of American agriculture kept American farm products cheap. Such mutually advantageous trade continued to grow as balanced improvements were concentrated in the British and American export fields; the rate of progress was even enough to insure a continuing gain to both sides. Under these conditions there was small probability of serious balance-of-payments strains.

The time came, however, when the rate of growth in productivity

---

4. Hicks does not discuss the possibility that *B* will shift to production of *A*-type goods, export these, and import *B*-export type goods, and that the new commodity terms of trade will be favorable to *B*. Nor does he deal with the question as to why should *B* not substitute *B*-export and *B*-domestic for *A*-export commodities. The probability that the rise in A-real incomes would tend to increase *B* exports to *A* also is not considered.

in Britain fell behind that in America. This, in itself, must have caused monetary difficulties. These difficulties were manageable so long as the disparity in the rate of progress remained small, but they became more serious as the disparity became wider. Between World Wars I and II the rate of growth in productivity in the United States was more rapid than in the industrial countries of Europe, including Britain. Hicks states that in the first phase of the dollar shortage (i.e., in the early 1920's) the difficulties which arose were almost exclusively monetary. British industry found it difficult to compete with American industry because British wages (under the restored gold standard) remained fixed in terms of dollars; but the productivity of agriculture (in North America and elsewhere) was still increasing so rapidly that a moderate devaluation would have restored Britain's earning power without imposing more than a very temporary sacrifice upon the British consumer. It was therefore correct, he argues, to think of the British-American mechanism-of-adjustment problems in the early 1920's in monetary terms.

Since then, we are told, the British (and western European) situation has become more difficult "mainly" because the rate of growth in productivity in American agriculture has slowed down. In addition, improvements in American industry have been concentrated upon products which compete most closely with imports from Britain (and western Europe). Consequently, the monetary adjustments needed to keep trade in balance under contemporary conditions may require a fall in the dollar value of British wages in order to keep British manufactures competitive with American manufactures. Britain, moreover, can no longer count upon the real sacrifice involved being temporary or slight, for she can no longer rely upon the increased productivity of American (and other) agricultural suppliers. The situation therefore requires a fall in British wages in terms of British imports, or a rise in the prices of British imports relatively to British wages. And the problems involved in attempting to implement policies which lower the level of real wages render the process of international adjustment extremely difficult.

Britain, we are reminded, has for long been unable to pay for her raw materials and foodstuffs by exporting directly to the "dollar area." Hicks says that for this, the United States tariff must bear some part of the blame; but since Britain has encountered similar difficulties in selling manufactures in Canada as in the United States, "it is clear that the increasing productivity of American industry must be regarded as mainly responsible."[5]

5. Hicks, *op. cit.*, p. 132.

"It is true that the forms in which the dollar problem has actually appeared have constantly tempted us to regard it as something much less deep-rooted. In the nineteen-twenties it was entangled with war debts; in the nineteen-thirties with world depression and the flight of capital before Hitler; in the nineteen-forties with war damage, more war debts, and the Cold War in international trade. A special explanation of the difficulties which were being experienced by *European countries* in settling their dollar debts could always be found. But the continuance of the same consequence, the same dollar shortage, as the result of these various 'causes,' has by now become very striking. That there is some *general* influence underlying these particular manifestations can now no longer be doubted. It is hard to see that there is any other general force which would account for what has been happening than the disparity in growth of productivity which we have been discussing."[6]

V

The available evidence does not lend support to the view that disparity in the rate of growth in productivity has been the general force which provides a satisfactory explanation of "the long-run dollar

TABLE I
COMPOUNDED ANNUAL RATE OF GROWTH IN PRODUCTIVITY
PER MAN-HOUR IN MANUFACTURING, 1920–1938

| Country | Selected Years — In Percentages | | | |
| | 1 | 2 | 3 | 4 |
	1920–1938	1924–1928	1929–1933	1934–1938
United States[1]	3½	2¾	−1¾	7½
United Kingdom[2]	2½	2	− ¾	4
France[3]	3	n.a.	n.a.	3½
Germany[4]	2½	n.s.	2½	n.s.
Netherlands[5]	3½	2½	3	4½
Sweden[6]	3½	2¼	3	2¼
Canada[7]	2¾	2	¼	5
Japan[8]	3	6	4½	n.s.
Australia[9]	2¼	3	−2	2

[1]Based on Solomon Fabricant, *Employment in Manufacturing 1899–1937*, pp. 230, 331; Colin Clark, *Conditions of Economic Progress* (London: Macmillan and Co., Ltd., 1951), pp. 271, 275, 279.
[2]*Census of Production*, London Board of Trade, 1908–1948.
[3]Clark, *loc. cit.*
[4]*Statistisches Jahrbuch für das Reich*, Internationaler Teil (Survey) (Berlin, 1938), p. 58.
[5]*Ibid.* Data for Netherlands reduced by 6 per cent to exclude mining.
[6]Clark, *loc. cit.*
[7]*Canada Year Book*, Canada Dominion Bureau of Statistics, Department of Trade and Commerce, 1870–1939.
[8]*Statistisches Jahrbuch für das Reich, loc. cit.*
[9]Clark, *loc. cit.*

6. *Ibid.*, p. 131. The discussion runs in terms of general "dollar-shortage" and not in terms of British "dollar-shortage." These are, of course, logically quite different problems, for no one has ever disputed that in a three-area world, with *A* and *B* competitive as to exports, *A* can suffer from more rapid technological improvement in *B*.

problem," so far as mechanism-of-adjustment strains or balance-of-payments difficulties are concerned. The compounded annual rate of growth in productivity per man-hour in manufacturing during the period 1920–1938 appears to have been about the same in the United States and in most industrial countries of western Europe — with the *exception of Britain*. As can be seen from Table I, roughly the compounded annual rate of growth in the United States, the Netherlands, and Sweden was 3.5 per cent; in France it was around 3 per cent. The figure of 2.5 per cent for Germany and 3 per cent for Japan is probably an underestimate; it omits large-scale military expenditures. Compared with most western European industrial countries Britain definitely lagged behind. The compounded annual rate of growth in productivity per man-hour in British manufacturing during the same period was 2.5 per cent. Britain's experience lends itself to careful case study; it does not necessarily lend itself to generalization.[7]

During the depression of 1929–1933, the differential rates of growth in productivity per man-hour in manufacturing actually worked in an equilibrating direction with respect to the international mechanism of adjustment. The United States suffered from a negative compounded annual rate of growth of −1.75 per cent, a lower figure than that of any western European country. The higher rates of growth in productivity abroad must have worked in the direction of lowering export prices of those countries and thereby improving their competitive position. For obvious reasons this was not a

7. Case studies dealing with Britain's economic development seem to confirm this conclusion. Our data with respect to Britain's comparative productivity growth have been checked against the information contained in the following studies: Walter Hoffmann, "The Growth of Industrial Production in Great Britain: A Quantitative Study," *Economic History Review*, Second series II (1949), 162–80; *idem, Wachstum und Wachstumformen der Englischen Industriewirtschaft von 1700 bis zur Gegenwart* (Jena: Gustav Fisher, 1940), especially pp. 261–84 for a useful series of comprehensive summary tables and graphs; L. Rostas, *Comparative Productivity in British and American Industry*, pp. 27–49, 76–93, 97–248; A. E. Kahn, *Great Britain in the World Economy*, particularly pp. 141–42, 185; Werner Schlote, *Entwicklung und Strukturwandlungen des englischen Aussenhandels von 1700 bis zur Gegenwart* (Jena: Gustav Fisher, 1938). Cf. also, Simon Kuznets, *Economic Change* (New York: W. W. Norton, 1953), pp. 253–77. Relative growth in productivity per labor unit is, of course, not decisive unless labor is the only cost, but it does provide a first approximation which appears to be confirmed by the other quantitative and qualitative information contained in these studies and by the information cited below. It must be kept in mind, however, that there are obvious limitations of product-per-labor-unit data as criteria of trends of double-factoral terms of trade, to say nothing of competitive ability. Aside from quality and design changes, shifts in tastes, etc., there is the whole field of cost or input items not covered by wages or labor-unit costs.

powerful factor at the time.   But it is noteworthy that while the rate of growth in manufacturing in the Netherlands and Sweden was 3 per cent, in Germany 2.5 per cent, and in Japan 4.5 per cent, that of Britain was −0.75 per cent.   From 1934 to 1938, on the other hand, the compounded annual rate of growth in productivity per man-hour in American manufacturing was 7.5 per cent, in Canada 5 per cent, in the Netherlands 4.5 per cent, in Britain 4 per cent, and in France 3.5 per cent.   This disparity in the rate of productivity growth in manufacturing occurred at the very time that the United States was moving toward equilibrium in its international accounts!

If differential rates of growth in over-all manufacturing productivity do not appear to have been an important general cause of long-term international imbalance, were improvements (as Hicks maintains) so concentrated in export industries on both sides of the Atlantic that this was the chief cause of the difficulty?   While detailed research on this problem is long overdue, and while it is impossible on the basis of the available information to reach a categorical conclusion, our analysis of the evidence suggests a partially affirmative but mainly negative reply.

The hypotheses seem plausible that countries are most likely to make their improvements in the production of products which they already make relatively well, and that it is these things which will most readily find an export market.   Improvements, says Hicks, will therefore be concentrated in the export fields;[8] and in earlier periods this probably occurred in Britain.[9]   In time, one would expect that some younger countries which were formerly customers of the older ones, but which possess the necessary natural resources for home production of the things they used to import, will acquire the necessary skill, adapt techniques to suit their needs, organize production

8. Hicks, *op. cit.*, p. 129.
9. Hicks presents his argument in a general form.   Economic historians appear to agree with it in so far as it applies to Britain.   Cf. A. D. Gayer, W. W. Rostow, A. J. Schwartz, and I. Frank, *The Growth and Fluctuation of the British Economy* (London: Oxford Clarendon Press, 1953), II, 647–58; W. W. Rostow, *British Economy of the Nineteenth Century* (Oxford: Oxford University Press, 1948), chaps. I and II; T. S. Ashton, *The Industrial Revolution 1760–1830* (Oxford: Oxford University Press, 1948), especially the discussion of interdependent factors in the fall of real costs; *idem*, "The Standard of Life of the Workers in England, 1790–1830," *The Tasks of Economic History*, Supplement IX, *Journal of Economic History* (1949), pp. 25–28.   For other countries one must consider the reverse hypothesis: progress may be most rapid in more backward industries (and countries) once they begin to import accumulated know-how from most advanced countries, since the latter can progress only through innovations (i.e., they cannot profitably imitate).

according to new methods, and make improvements in industries which compete closely with the imports from the more developed countries. To some extent this certainly happened in the United States. And it is of course true that in the latter half of the nineteenth century vast areas in the United States were brought into

TABLE II

INCREASE IN THE RATIO OF FARM OUTPUTS TO TOTAL INPUTS, UNITED STATES, 1910–1950

(1910–1914 = 100)

Year	Farm output per unit of input[1]	Year	Farm output per unit of input[1]
1910	100	1931	117
1911	98	1932	117
1912	106	1933	111
1913	95	1934	99
1914	101	1935	119
1915	102	1936	105
1916	96	1937	130
1917	104	1938	125
1918	101	1939	126
1919	99	1940	130
1920	102	1941	138
1921	92	1942	152
1922	102	1943	147
1923	104	1944	151
1924	106	1945	149
1925	108	1946	153
1926	111	1947	146
1927	108	1948	158
1928	112	1949	153
1929	111	1950	153
1930	106		

Source: T. W. Schultz, *The Economic Organization of Agriculture*, p. 122.
[1]Inputs weighted by 1946–1948 prices.

cultivation and became a source of cheap agricultural exports. But is it true, as Hicks contends, that before the 1920's American improvements were concentrated mainly in the production of agricultural exports and since then improvements have seriously lagged behind in the production of these products? Is there any evidence to show that generally speaking improvements in the United States have not been quite evenly spread among the production of domestic, export, and import substitute products?

Excellent work has recently been done in attempts to measure

the rate of change in productivity in American agriculture.[1] The evidence reveals remarkable progress, rather than decline in the rate of productivity growth between 1920 and 1950. The increase in the ratio of farm outputs per unit of input during this period is presented in Table II. Taking 1946–1948 input prices for weighting the various inputs, one finds that a unit of input in 1950 resulted in 53 per cent more agricultural production than in 1910. In 1950 (or 1948) the output of American agricultural production was 75 per cent larger than in 1910, and only 14 per cent more inputs were required to produce this expansion.[2] Virtually all the improvement took place since 1923; and when it is averaged over the twenty-seven year period 1923–1950, it represents an average increase of 2 per cent per year.[3]

The contention is unwarranted that from the early 1920's (until the early 1950's) there was a decline in the rate of productivity in American agriculture, and that this was an important factor contributing to the long-term dollar shortage.

Whether or not improvements in American industry were concentrated upon the production of products which compete most closely with European *imports* is a problem more difficult to resolve. Goods that might be classified as domestic, import-substitute, or export products were continuously undergoing change. Without specific knowledge to the contrary, one would expect that owing to the well-balanced resource structure of the United States, and the greater importance of the domestic as compared with the foreign market, improvements would have been distributed quite uniformly as between the production of domestic, export, and import-substitute products. Within each one of these categories, however, one would expect some industries to lead in the rate of growth, then to slacken

1. Cf. Theodore W. Schultz, *The Economic Organization of Agriculture* (New York: McGraw-Hill Book Co., 1953), chaps. 7 and 8.

2. *Ibid.*, p. 120.

3. Schultz, *op. cit.*, pp. 99–145. Cf. also, Bureau of Agricultural Economics, *Farm Income Situation*, July–Sept., 1951, Tables 1, 11, 12. New and better production techniques were mainly responsible for the advance in efficiency. Labor was withdrawn and cheaper inputs were added. Farm production expenses on the operation of vehicles (less hired labor) amounted to about $7 million in 1910, and $2,073 million in 1950, a 296-fold increase; seed purchased increased 10 fold; fertilizer and lime, 551 per cent. It has been estimated that the federal and state governments allocated about $32 million for agricultural research in 1937 and about $106 million in 1951. This progress in research and the resulting increase in the production of new techniques was quite continuous. The returns to our society were very large. Land did not become a limitational factor, responsible for historical diminishing returns. In effect, it was demonstrated that there is no general *historical* "law" of diminishing returns to inputs used on agricultural land, so long as there is persistent and large-scale technological progress.

TABLE III

AVERAGE RATE OF PERCENTAGE INCREASE DURING FIVE-YEAR PERIODS, SHOWN BY INDICES OF THE VOLUME OF PRODUCTIVE ACTIVITY IN VARIOUS INDUSTRIES, IN THE UNITED STATES, GREAT BRITAIN, BELGIUM, GERMANY AND FRANCE

Country and Nature of Series	Period Covered	Average Rate of % Increase During 1st Half	During 2nd Half	Relative Decrease in Rate (3 −4 as % of 3)
1	2	3	4	5
United States				
1. Wheat Crops.............	1866–1924	20.4	7.3	64.2
2. Corn Crops.............	1866–1924	17.7	6.7	62.1
3. Potato Crops...........	1866–1924	17.6	11.5	34.7
4. Cotton Crops...........	1866–1924	24.2	5.6	76.9
5. Anthracite Coal Shipments	1825–1924	70.8	16.3	77.0
6. Bituminous Coal Output..	1840–1924	62.9	33.9	46.1
7. Crude Petroleum Output..	1860–1924	70.5	51.4	27.1
8. Pig Iron Production......	1856–1924	46.7	29.4	37.0
9. Crude Steel Production...	1866–1924	157.7	46.2	70.7
10. Portland Cement Output..	1880–1924	337.4	90.0	26.7
11. Cotton Consumption, Domestic Mills........	1871–1924	30.4	19.0	37.5
12. Raw Silk Imports........	1866–1924	66.0	41.8	63.3
13. Locomotives Produced, Baldwin Locomotive Works...............	1836–1923	47.9	22.3	53.4
Great Britain				
14. Coal Output.............	1856–1913	16.2	10.4	35.8
15. Pig Iron Production......	1856–1913	15.4	4.0	74.0
16. Raw Steel Output........	1876–1913	39.3	20.4	48.1
17. Tonnage of Ships Cleared, All Ports..............	1816–1913	24.8	18.9	23.8
18. Raw Cotton Imports.....	1781–1913	35.4	11.1	68.6
19. Tea Consumption[1].......	1810–1919	14.8	12.6	14.9
Belgium				
20. Coal Output.............	1831–1913	25.6	6.9	73.0
21. Pig Iron Output.........	1851–1913	18.8	21.6	14.9
22. Steel (Crude) Output.....	1881–1913	58.9	46.8	20.5
23. Zinc Production.........	1846–1913	29.0	14.5	50.0
Germany				
24. Wheat Crops............	1881–1913	14.7	6.7	54.4
25. Coal Output.............	1861–1913	28.6	25.7	10.1
26. Pig Iron Consumption....	1861–1913	36.7	25.5	30.5
27. Steel Output............	1881–1913	73.8	67.8	8.1
28. Zinc (Crude) Production..	1846–1913	24.7	10.5	57.5
29. Raw Cotton Consumption[2]	1836–1910	43.5	22.8	47.6
France				
30. Wheat Crops............	1825–1913	7.8	2.0	74.4
31. Coal Output.............	1811–1913	29.6	15.9	46.3
32. Petroleum Consumption...	1866–1913	57.9	26.6	54.1
33. Pig Iron Output.........	1825–1913	24.3	16.8	30.9
34. Steel Output............	1871–1913	47.3	35.4	25.2

Source: Simon Kuznets, *Economic Change*, p. 256. The results were obtained by taking the volume of output or consumption for an industry, averaging it for nine years (or seven where the data were short at the end of the series) centered around the year ending with 0 or 5 and dividing one average by the preceding. The percentage rate of increase or decrease by five-year periods was thus obtained. To see whether the general tendency was to decline, rise, or remain stable, the entire period covered by each series was divided into halves, and an average percentage rate of increase for each half obtained. Of thirty-five series only one showed an increase in the average percentage rate of increase.

[1] Ten-year averages until 1854.
[2] Ten-year averages through the whole period.

(occasionally to take a further spurt), and, in time, to be overtaken by others whose period of rapid growth was just beginning. The available evidence, inadequate and tentative as it may be, appears to confirm this hypothesis. Table III summarizes the average rate of percentage increase during five-year periods of the volume of productive activity in various industries (domestic, export, and import substitute) in the United States, Great Britain, Belgium, Germany, and France.

During the first half of the periods covered (see Table III), the rate of growth in activity in selected industries in the United States was generally much higher than in Britain. But during the second half of the periods covered the rate of expansion in the United States industries was greatly reduced as compared with the other countries.[4] During both the first and the second half of the periods covered, with the exception of the remarkable rates of growth in the output of cement and steel, the average percentage rate of increase in the volume of productive activity in domestic and export industries — such as bituminous coal, pig iron, crude petroleum, anthracite coal, cotton, and wheat were not dissimilar to the average rate of percentage increase in the import-competing industries — such as cotton and silk manufactures. While variation in output change exceeded variation in efficiency change, virtually every industry in the United States showed an upward trend in efficiency.[5]

As can be seen from Table III, even before World War I, Britain was unable to maintain her rate of growth in output of such industries as coal, iron, and steel, as compared with the rate of growth either in

4. The spectacular rates of growth between 1866 and 1888 and the very high rates between 1894 and 1914 help to explain the much higher level of per capita real income the United States attained, as compared with the industrial countries of Europe, before World War I, and has maintained at a compound (though irregular) rate thereafter. As already observed, between World Wars I and II the rate of growth per man-hour in manufacturing was not greatly dissimilar in the United States and in most industrial countries of Europe. Professor Kuznets' data show that from 1869 to 1888 the United States national income increased at a rate of almost 39 per cent per quinquennium, and per capita income increased at a rate of about 23 per cent per quinquennium. From 1869 to 1938 the national income increased at a rate of about 19 per cent per quinquennium while the per capita income increased at an 8.5 per cent rate. Computed from Simon Kuznets, *National Income: A Summary and Findings* (New York: National Bureau of Economic Research, 1946), Table 10, p. 32.

5. For an informative survey cf. Solomon Fabricant, *Economic Progress and Economic Change* (New York: National Bureau of Economic Research, 34th Annual Report, 1954), pp. 3–18, and the list of relevant National Bureau publications on pp. 85–88. A good selected bibliography may be found in L. Rostas. *op. cit.*, pp. 249–59.

the United States or in Belgium, Germany, and France. She was lagging behind practically all the rapidly industrializing countries of the world. World War I and its aftermath did not improve this trend. We have observed that during the period 1920 to 1938 the compounded annual rate of growth in productivity per man-hour in *manufacturing* in Britain amounted to about 2.5 per cent while in the United States it was about 3.5 per cent.[6] In British manufacturing, mining, building, and public utilities, over the thirty-year period from 1907 to 1937, the increase in output per man-hour amounted to 65 per cent and in output per wage-earner to 47 per cent.[7] In the same period the increase in the United States amounted to 133 per cent in output per man-hour and to 71 per cent in output per wage earner.[8] The rate of increase in productivity in terms of man-hours was thus nearly twice as high in the United States as in Britain. Over the thirty-year period, the average compound rate of increase amounted to 1.7 per cent per man-hour per year in Britain, and to 2.9 per cent per man-hour per year in the United States.[9]

Similarly, between 1908 and 1936–1937 output increased somewhat over 40 per cent per man-hour and about 30 per cent per head in British agriculture; in the comparable period of 1909 to 1937 the increase of productivity in American agriculture amounted to about 66 per cent per man-hour and to about 58 per cent per head.[1] The rate of increase per head per year was a little over 2 per cent in the United States, and only 1 per cent in Britain.[2] Between 1937 and 1944 there was a further 18 per cent increase in output per head (and perhaps a little less per man-hour) in American agriculture, while in Britain the wartime increase per man-hour has been estimated at 13 per cent.[3]

Data relating to productivity comparisons in individual industries reveal the same tendency, with practically all the British export industries lagging behind comparable industries in the United States. From 1909 to 1939 British output per man-hour in the cotton industry increased by only 46 per cent, while in the United States output per man-hour more than doubled;[4] in the woolen industry British output

6. Cf. Table I.
7. L. Rostas, *op. cit.*, p. 42.
8. *Loc. cit.*
9. The results are summarized in Table IV.
1. L. Rostas, *op. cit.*, p. 79.
2. *Loc. cit.*
3. *Loc. cit.*
4. *Ibid.*, Appendix 9, p. 136.

per man-hour increased 26–35 per cent, while in the United States output per man-hour increased 78 per cent.[5] The rate of increase was substantially greater in United States steel rolling, vehicles, chemicals, mining, tobacco, clothing, food and drink, etc. There was in fact no single main industry group in which the rate of increase in output per man-hour in the United States was smaller than in Britain.[6]

TABLE IV

AVERAGE COMPOUND PERCENTAGE RATE OF ANNUAL INCREASE IN PER MAN-HOUR PRODUCTIVITY IN UNITED KINGDOM AND UNITED STATES INDUSTRY AS A WHOLE, 1907–1937

(1907 = 100)

	United Kingdom	United States
1907–1937.........	1.7	2.9
1909–1924.........	1.4	2.8
1924–1937.........	2.1	3.0

Source: Adapted from L. Rostas, *Comparative Productivity in British and American Industry*, p. 42.

VI

However, the rate of increase in output in the United States as compared with Britain between World Wars I and II does not suggest that improvements were concentrated in American import-competing industries. If one compares United States output per worker with British output per worker in similar industries before World War II, one gets the impression that not only has improvement in American efficiency tended to rise persistently, but it has also tended to rise in all sectors of the economy. The ratio of United States to British output per worker in different industries, and the corresponding ratio of United States to British exports is presented in Table V.

If, as one would expect, industrial improvements in the United States and in Britain were reflected in the relative output per worker in comparable industries, then the fact that United States output per worker before World War II was greater than the British in domestic, export, and import-substitute industries suggests that American improvements were not concentrated particularly in the import-competing fields. United States output per worker was more than

5. *Ibid.*, Appendix 10, p. 144.
6. *Ibid.*, p. 46; Appendixes 1–37, pp. 97–248.

TABLE V

RATIO OF UNITED STATES TO BRITISH OUTPUT PER WORKER AND OF
UNITED STATES TO BRITISH EXPORTS IN SELECTED INDUSTRIES[1]

Industry	Output per worker U. S. : U. K.	Exports U. S. : U. K.
Electric lamps	5.4	0.94
Tin cans	5.25	3.0
Pig Iron	3.6	5.1
Wireless receiving sets and valves	3.5	7.6
Motor cars	3.1	4.3
Biscuits	3.1	0.23
Matches	3.1	0.09
Rubber Tires	2.7	0.74
Soap	2.7	0.35
Machinery	2.7	1.5
Glass containers	2.4	3.5
Paper	2.2	1.0
Beer	2.0	0.056
Linoleum, oilcloth, etc.	1.9	0.34
Coke	1.9	0.19
Hosiery	1.8	0.30
Cigarettes	1.7	0.47
Rayon weaving	1.5	0.20
Cotton spinning and weaving	1.5	0.11
Leather footwear	1.4	0.32
Rayon making	1.4	0.091
Woolen and Worsted	1.35	0.004
Men's and boys' outer clothing, wool	1.25	0.044
Margarine	1.2	0.031
Cement	1.1	0.091

Source: L. Rostas, *op. cit.;* G. D. A. MacDougall, "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part 1," *Economic Journal,* LXI (Dec. 1951), 700. Before World War 11, American weekly wages in manufacturing were roughly double the British. MacDougall has shown that where American output per worker was more than twice the British, the United States has in general the bulk of the export market (electric lamps, biscuits matches, rubber tires and soap, being exceptions), while for products where it was less than twice as high the bulk of the market was held by Britain. *Op. cit.,* p. 698.

[1]1935, 1937 or 1939 in each country with minor exceptions. Rostas suggests the following rough relative productivities per worker, U. S. : U. K., in certain nonmanufacturing sectors: mining, 4.15; communications, 2.7; electricity, 1.9; gas, 1.7; distribution, 1.5; building, 1.15; commercial and other services, 1.0; transport of goods, 1.0. *Op. cit.,* p. 89. Cf. MacDougall, p. 709.

twice the British in the production of such products as electric lamps, mining, tin cans, pig iron, radios, motor cars, biscuits, matches, rubber tires, soap, machinery, glass containers, and paper. It was less than or equal to 1.5 times the British in the production of woolens and worsted, cotton spinning and weaving, rayon weaving, men's and boys' outer clothing of wool, and leather footwear. It was about the same in the production of cement, building, commercial services, and

transport.[7] The figures suggest that American efficiency has been relatively greater than the British in the capital-intensive (and other) industries which have been able to take especial advantage of mass scale production and distribution methods at *home*, and which have probably also benefited from enlarged markets abroad. Many of these industries have competed successfully within the British market and in third countries; but they surely cannot be differentiated as United States import-competing industries. Those that can (e.g., woolens and worsted, cotton spinning and weaving) manifestly were not ahead of most other American industries in efficiency growth. In the strictly domestic fields (e.g., building and commercial services), American and British productivity appear to have been about the same. Apparently British balance-of-payments difficulties cannot be attributed primarily to a tendency for American improvements to be concentrated in import-competing industries.[8] Britain generally lagged behind in productivity growth not only as compared with the United States, but also as compared with the leading industrial countries of western Europe.

VII

Britain's declining position in world trade was doubtless interrelated with the aforementioned widening gaps in industrial efficiency. It is therefore not surprising to find that explanations of long-term international imbalance which are expressed in terms of differential rates of change in productivity are usually "reinforced" with material dealing with "the changing structure of world trade." It appears doubtful, however, whether the long-term imbalance of Britain — and particularly any such tendency of western Europe as a whole — vis-à-vis America can be explained in such terms.

From 1899 to 1937 Britain's relative share in world trade decreased by a greater amount than that of any other country. It dropped from 32.5 per cent to 22.4 per cent.[9] Britain's exports to North America amounted to 10.3 per cent of her total exports in 1900, while her exports to Europe amounted to 41 per cent. In 1910 her exports to North America amounted to 12.7 per cent and those

7. Cf. Table V.
8. For data on the extent of American capital-intensive type of imports see Wassily Leontief, "Domestic Production and Foreign Trade; The American Capital Position Re-Examined," *Proceedings of the American Philosophical Society*, Vol. 97, No. 4 (Sept. 1953), pp. 332–49.
9. Cf. Table VI.

to Europe 34.9 per cent. In 1929 the ratios were 11.7 per cent and 34.7 per cent respectively. By 1936 British exports to North America still amounted to 11.9 per cent of her total exports and those to Europe to 35.8 per cent.[1] Similarly, the long-term trend shows that Britain's export position deteriorated in terms of all industrial countries and not only in terms of North America. British exports of finished manufactured goods to the United States rose steadily from about £5.5 million in 1827–1830 to £31.9 million in 1927–1929. But of her total exports to the United States, 93.2 per cent were finished manufactured goods in 1827–1830, whereas only 67.5 per cent were finished manufactured goods in 1909–1913, a decline of 25.7 percentage points. Of her exports to "industrial Europe," finished manufactures amounted to 83.4 per cent in 1827–1830, and 60.5 per cent in 1909–1913, a decline of about 22.9 percentage points. British exports of finished manufactured goods to all industrial countries amounted to 86.9 per cent of her total exports in 1827–1830 and 61.9 per cent in 1909–1913.[2] None of the other industrial countries of Europe fared as badly. Indeed, from 1899 to 1937 the percentage shares in world trade of Sweden, Belgium, and Germany actually rose somewhat, those of Switzerland and Italy fell slightly, but that of Britain fell 10.1 and that of France 9.4 percentage points. Canada, on the other hand, gained 4.7, Japan 5.7, and the United States 8.4 percentage points.[3]

VIII

The drastic decline in Britain's share of world trade cannot be explained directly in terms of the changing structure of world trade, if by that is meant her failure to concentrate on expanding industries, for Britain lost ground in the stable and declining industries as well as in the expanding ones.[4] She was unable to keep pace with the manufactured products of the expanding industrial nations, and she was unable to retain her competitive position with respect to them in the stable and declining fields as well. It was not the case that Britain's share of world trade in each one of these industry-groups remained about the same, while the industries in which she concentrated her production declined in importance. By subtracting a country's share of world trade in the expanding, stable, and declining

1. Cf. W. Schlote, *British Overseas Trade from 1700 to the 1930's*, trans. W. O. Henderson and W. H. Chaloner (Oxford: Blackwell, 1953), pp. 159–60.
2. Cf. Table VII.
3. Cf. Table VI.
4. Cf. Table VIII.

TABLE VI

CHANGES IN WORLD SHARE AND STRUCTURE OF TRADE OF ELEVEN COUNTRIES
1899, 1937, 1950

		Share in World Trade	Structure of Trade				
			Expanding	Stable	Declining	N. C.	Total
U. S. A.	1899	11.2	23.4	54.1	21.7	0.8	100
	1937	19.6	58.8	32.4	8.5	0.3	100
	Change	8.4	35.4	−21.7	−13.2	−0.5	0
	1950	29.1	52.4	32.0	13.4	2.0	100
	Change	9.5	− 6.4	− 0.4	4.9	1.7	0
United Kingdom	1899	32.5	17.7	18.6	62.9	0.8	100
	1937	22.4	31.2	25.0	43.6	0.2	100
	Change	−10.1	13.5	6.4	−19.3	−0.6	0
	1950	25.0	43.0	24.8	31.6	0.4	100
	Change	2.6	11.8	− 0.2	−12.0	0.2	0
France	1899	15.8	5.8	31.0	62.8	0.4	100
	1937	6.4	27.9	36.1	35.8	0.2	100
	Change	− 9.4	22.1	5.1	−27.0	−0.2	0
	1950	10.2	34.0	27.4	38.3	0.6	100
	Change	3.8	6.1	− 8.7	2.5	0.4	0
Germany	1899	22.2	11.2	45.8	42.5	0.3	100
	1937	22.4	35.29	46.14	18.44	0.10	100
	Change	0.2	24.09	0.34	−24.06	0.1	0
	1950	7.1	43.5	44.3	12.1	0.3	100
	Change	−15.3	8.21	− 1.84	− 6.34	−0.07	0
Belgium	1899	5.6	13.64	49.21	35.48	1.65	100
	1937	5.9	35.01	41.26	23.71	0.004	100
	Change	0.3	21.37	− 7.95	−11.77	−1.646	0
	1950	5.8	33.6	37.1	29.2	—	100
	Change	− 0.1	− 1.41	− 4.16	5.49	− 0.004	0
Sweden	1899	1.0	46.24	45.39	8.12	0.23	100
	1937	2.5	49.19	40.80	9.98	0.015	100
	Change	1.5	2.95	− 4.59	1.86	− 0.215	0
	1950	2.7	42.1	40.80	15.9	1.1	100
	Change	0.2	− 7.09	—	5.92	1.0	0
Canada	1899	0.3	9.3	57.8	22.6	10.3	100
	1937	5.0	14.9	75.4	9.6	0.1	100
	Change	4.7	5.6	17.6	−13.0	−10.2	0
	1950	6.1	11.2	80.3	8.1	0.2	100
	Change	1.1	− 3.7	4.9	− 1.5	0.1	0
Switzerland	1899	3.9	73.60	25.88	0.6	—	100
	1937	2.9	21.89	50.58	27.5	—	100
	Change	− 1.0	−51.71	24.70	26.9	—	0
	1950	4.0	29.80	50.8	19.2	—	100
	Change	1.1	7.91	0.22	− 8.3	—	0

TABLE VI (*Continued*)

CHANGES IN WORLD SHARE AND STRUCTURE OF TRADE OF ELEVEN COUNTRIES
1899, 1937, 1950

		Share in World Trade	Structure of Trade				
			Expanding	Stable	Declining	N. C.	Total
Japan	1899	1.5	0.1	25.3	72.8	1.8	100
	1937	7.2	13.3	18.4	68.0	0.3	100
	Change	5.7	13.2	− 6.9	− 4.8	− 1.5	0
	1950	3.3	16.62	21.85	61.36	0.15	100
	Change	− 3.9	3.32	3.45	− 6.64	− 0.15	0
Italy	1899	3.7	1.64	14.71	83.64	—	100
	1937	3.6	23.64	17.62	58.72	—	100
	Change	− 0.1	22.0	2.91	−24.92	—	0
	1950	3.8	26.1	15.6	58.1	0.2	100
	Change	0.2	2.46	− 2.02	− 0.62	0.2	0
India	1899	2.3	0.006	30.85	68.88	0.25	100
	1937	2.1	5.19	22.15	72.22	0.41	100
	Change	− 0.2	5.184	− 8.70	3.34	0.16	0
	1950	2.9	0.6	9.3	89.7	0.4	100
	Change	0.8	− 4.59	−12.85	17.48	− 0.01	0
World	1899	—	13.4	33.2	52.8	0.6	100
	1937	—	34.8	35.7	29.3	0.2	100
	Change	—	21.4	2.5	−23.5	− 0.4	0
	1950	—	39.1	33.3	26.8	0.8	100
	Change	—	4.3	− 2.4	− 2.5	0.6	0

Source: Calculated from data presented in H. Tyszynski, "World Trade in Manufactured Commodities, 1899–1950," *Manchester School of Economics and Social Studies*, XIX (Sept. 1951), 277–81.

TABLE VII

EXPORTS OF BRITAIN'S HOME PRODUCTS TO INDUSTRIAL COUNTRIES BY COMMODITY GROUPS, 1827–1929

Commodity Group	In £1,000,000						Per cent					
	1827–1830	1854–1857	1877–1879	1898–1901	1909–1913	1927–1929[1]	1827–1830	1854–1857	1877–1879	1898–1901	1909–1913	1927–1929[1]
Exports to Industrial Europe												
Foodstuffs................	1.0	1.6	2.8	3.0	5.3	6.9	10.6	5.8	5.1	4.2	4.9	5.2
Raw materials............	0.2	3.9	7.3	19.3	27.8	46.3	2.3	14.2	13.3	27.7	25.5	34.7
Finished manufactured goods.....	7.8	20.3	41.2	42.4	65.8	80.3	83.4	74.7	74.8	60.9	60.5	60.1
Unclassified..............	0.3	1.3	3.7	4.9	10.0	—	3.7	5.3	6.8	7.2	9.1	—
Total...............	9.3	27.1	55.0	69.6	108.9	133.5	100	100	100	100	100	100
Exports to the United States:												
Foodstuffs................	—	0.3	0.1	0.7	2.2	1.9	—	1.5	0.6	3.9	7.5	4.1
Raw materials............	—	1.1	2.1	2.2	5.5	12.1	—	5.5	12.3	12.4	18.6	26.4
Finished manufactured goods.....	5.5	17.9	14.2	14.1	20.0	31.9	93.2	89.9	83.0	79.2	67.5	69.5
Unclassified..............	0.4	0.6	0.7	0.8	1.9	—	6.8	5.1	4.1	4.5	6.4	—
Total...............	5.9	19.9	17.1	17.8	29.6	45.9	100	100	100	100	100	100
Exports to All Industrial Countries:												
Finished manufactured goods......	13.3	38.2	55.4	56.5	85.8	112.2	86.9	81.1	76.9	64.6	61.9	62.5
Raw materials and foodstuffs[1].....	1.9	8.8	16.7	30.9	52.7	67.2	13.1	18.9	23.1	35.4	38.1	37.5
Total...............	—	—	—	—	—	—	100	100	100	100	100	100
Exports to All Agrarian Countries:												
Finished manufactured goods.....	20.3	54.6	111.3	149.7	259.8	450.8	93.3	90.0	91.0	83.2	82.1	83.3
Total Exports...............	37.0	107.7	194.4	267.3	455.0	720.7						

Source: W. Schlote, *Entwicklung und Strukturwandlungen des englischen Aussenhandels von 1700 bis zur Gegenwart, op. cit.*, p. 92.
[1]Excluding Eire.

TABLE VIII

Change in Share of World Trade in Expanding, Stable and Declining Industry Groups of Eleven Countries, 1899–1937

	By Group No.	U.K.	U.S.A.	France	Germany	Belgium	Italy	Sweden	Switzerland	Canada	India	Japan
Expanding Groups												
Motor vehicles, etc.	10	− 6.6	22.5	−15.8	− 9.4	− 2.7	6.1	0.6	−0.4	3.2	—	1.8
Industrial Equipment	6	−18.1	7.2	− 4.5	10.2	3.7	3.8	−0.2	0.8	0.8	−0.08	2.7
Iron and Steel	1	−29.4	9.5	3.0	4.3	7.8	0.5	−1.3	0.09	1.0	1.1	3.2
Electric Goods	7	−10.7	− 3.0	− 2.1	10.7	− 3.4	−0.6	3.8	−7.1	1.9	—	3.8
Total Expanding Groups		−22.9	13.6	− 1.7	4.0	0.3	2.0	0.2	−0.3	1.9	0.3	2.7
Stable Groups												
Miscellaneous Material	5	2.1	− 6.2	−11.9	−11.6	− 1.1	−0.7	5.5	−0.06	22.7	−1.8	3.4
Non-ferrous metals	2	− 2.1	−20.6	− 8.5	− 6.8	6.5	0.3	1.1	2.4	28.4	2.0	2.6
Chemicals	3	− 5.4	4.1	− 2.2	3.1	− 3.1	0.02	0.3	2.9	2.4	−1.1	1.6
Agricultural Equipment	8	−22.7	2.1	− 1.2	15.4	0.5	0.5	−1.5	−0.2	2.7	—	0.3
Met. Manufacturers, NCS	14	− 6.8	− 1.2	− 7.4	9.6	− 2.5	0.8	1.4	0.5	1.3	−0.02	4.2
Books, Films, etc.	15	6.9	13.9	−16.9	−13.9	1.5	−0.2	0.3	−0.07	1.6	0.2	6.8
Non-metalliferous	4	4.4	8.0	−10.4	5.0	−13.1	1.7	−2.0	0.2	0.3	0.13	5.8
Total Stable Groups		− 2.6	− 0.4	− 8.2	− 1.8	− 1.4	0.1	1.5	1.0	10.0	−0.8	2.6
Declining Groups												
Drinks and Tobacco	11	26.9	− 2.8	−22.2	− 7.1	5.4	−3.8	0.04	−0.3	8.6	−0.15	1.4
Railways, Ships, etc.	9	−27.9	− 3.4	0.3	19.3	− 1.9	0.3	6.9	−0.2	0.6	−0.03	5.9
Miscellaneous Mfg.	16	3.8	2.8	−11.1	− 5.2	− 0.8	−0.04	−0.5	−0.3	0.4	0.12	11.0
Apparel	13	− 1.6	5.4	−27.1	−10.4	1.7	3.7	0.5	1.6	4.7	0.2	21.3
Textiles	12	− 8.8	1.0	− 8.2	− 4.2	3.9	1.6	0.2	−0.4	0.2	3.42	16.8
Total Declining Groups		− 5.4	1.0	−11.0	− 3.8	1.0	1.3	0.7	−2.3	1.5	−2.3	14.7

Source: Tyszynski, *op. cit.* Calculated from Tables I–V, pp. 276–82. Cf. also Table XI, p. 290.

TABLE IX

CHANGE IN SHARE OF WORLD TRADE IN EXPANDING, STABLE AND DECLINING INDUSTRY GROUPS OF ELEVEN COUNTRIES, 1937–1950

	By Group No.	U.K.	U.S.A.	France	Germany[1]	Belgium	Italy	Sweden	Switzerland	Canada	India	Japan
Expanding Groups												
Motor vehicles, etc.....	10	17.0	−2.8	2.6	−7.4	−1.1	−3.9	−0.1	−0.3	2.6	−6.0	−1.4
Industrial Equipment....	6	4.1	11.3	2.9	−21.6	0.6	−1.0	3.2	2.5	0.1	−3.3	−1.9
Iron and Steel....	1	3.6	2.3	8.4	−8.2	0.6	0.7	−2.0	0.06	1.6	−0.7	−1.2
Electric Goods....	7	11.2	7.3	4.8	−23.8	0.6	1.6	−0.5	2.3	0.5	0	−2.9
Total Expanding Groups..		7.4	5.8	3.1	−14.8	−1.0	0	−0.4	1.3	−0.4	0.1	−1.3
Stable Groups												
Miscellaneous Material..	5	−0.5	5.3	1.2	−17.9	−1.9	−0.3	2.7	0.3	15.8	2.0	3.2
Non-ferrous metals......	2	−4.2	9.3	1.3	−1.7	−0.7	0.2	−0.1	2.9	0.8	−0.8	4.6
Chemicals....	3	0.6	20.1	0.2	−24.3	1.7	−0.5	0.2	2.2	2.0	−1.9	2.4
Agricultural Equipment..	8	7.1	2.6	1.1	−8.5	−0.2	0.7	−1.5	0.05	4.3	0	0.26
Met. Manufacturers, NCS	14	7.0	10.6	5.5	−29.1	0.5	0.7	0.4	6.5	0.4	−1.1	−1.8
Books, Films, etc....	15	3.0	12.4	2.2	−19.0	1.4	0.8	0.2	2.3	0.7	−2.0	2.6
Non-metalliferous......	4	5.9	10.4	5.7	−21.3	−4.4	−1.3	−0.32	0	8.4	−3.3	3.1
Total Stable Groups......		3.0	10.2	2.0	−19.4	−0.4	0	0.5	2.1	4.1	−0.5	−1.5
Declining Groups												
Drinks and Tobacco....	11	−9.2	7.4	5.6	−3.1	0.4	−4.2	0.16	2.5	1.2	−8.3	−1.7
Railways, Ships, etc....	9	−3.7	13.4	3.4	−25.8	−2.6	2.7	2.8	0.5	4.2	−1.3	−2.2
Miscellaneous Mfg....	16	14.3	7.2	6.4	−24.8	0	−0.1	0.6	2.4	0.3	−3.9	−6.2
Apparel....	13	7.9	16.0	10.6	−20.0	2.7	0	0.3	1.0	3.2	−6.0	−16.1
Textiles....	12	−8.4	7.2	7.5	−6.6	−0.1	1.8	−0.03	−0.3	0.1	6.6	−10.0
Total Declining Groups....		−3.8	5.8	6.8	−10.9	1.6	1.0	8.0	3.0	0.2	4.5	−9.1

Source: Calculated from data presented in Tyszynski, *op. cit.*, vol. xix, pp. 276–82.

[1] Western Control.

industries in 1899 from its share in 1937, one can estimate the share of world trade gained or lost by selected countries in each group of industries. This information is summarized in Table VIII. In the expanding industries Britain lost ground in each group. However, had she retained in 1937 the share of trade in the iron and steel and engineering industries that she held in 1899, her share of world trade would have fallen only from 32.5 per cent to 31.3 per cent, instead of to 22.4 per cent.[5] Each of the three countries which gained most in its relative importance in world trade — United States, Japan, and Canada — advanced by means of improving its competitive position in different groups of industries. The American advance was mainly in the expanding industries, that of Canada in the stable ones and that of Japan in the declining groups. Other countries, such as Germany, gained somewhat in the expanding industries; Belgium, in the declining and expanding ones; Italy, in all three groups; Switzerland, in the stable ones; Sweden and India, in the declining industries. France, on the other hand, lost ground in all three industry-groups, despite the fact that she experienced a sharp movement from the declining toward the expanding industries.

As can be seen from Table VIII, the United States forged ahead in the expanding fields; her share of world trade rose substantially in motor vehicles, in iron and steel, and in industrial equipment. Japan and Canada gained in each group of the expanding fields. Britain lost ground in each group. Germany made striking gains in industrial and electrical goods; Belgium in iron and steel; and Sweden in electrical goods. In the stable fields, the United States was unable to maintain her relative position, falling behind in nonferrous metals, but gaining in films. Canada improved her relative position most strikingly in these fields, with nonferrous metals and paper leading her advance. Japan also gained in virtually every one of the expanding groups, and Belgium increased her share substantially in nonferrous metals. Britain lost ground in the stable fields as a whole, and heavily in agricultural equipment — precisely the field in which Germany had the greatest gain. In the fields of declining importance in world trade, Japan and Canada made gains in every group. Japan's greatest gains were in apparel, textiles, and miscellaneous consumer goods. Britain lost ground heavily in the declining industries, particularly in railways and ships, and in textiles.

5. Cf. H. Tyszynski, "World Trade in Manufactured Commodities, 1899–) 1950," *Manchester School of Economics and Social Studies*, XIX (Sept. 1951, 291–92.

Because of Britain's early economic development, and of the predominant share of the declining industries in the structure of her trade and economy in the nineteenth century, adaptation to the changing pattern of world trade was made more difficult in all fields.[6] In some of the expanding industries Britain showed very rapid domestic growth. This was especially true during the 1920's and 1930's in power, automobiles, electrical machinery, rayon, and chemicals. The proportion of Britain's trade accounted for by the expanding industries increased from 17.7 per cent of her total trade in 1899 to 31.2 per cent in 1937.[7] While this was a significant shift, it was smaller than the shift of world trade as a whole in these industries.[8] Unfortunately the extent of the British adjustment was insufficient to relieve her from contraction of export income. Not only was there a world shift toward those industries in which British exports were at all times less important, but, in addition, there was a decreasing importance of exports in each individual British industry.[9] The decline in exports was quite general; it revealed, in part, the world's greater competitive power as compared with that of Britain. The rapid rise of British imports of the new manufactured products told the same story. The underlying trade position was moving against Britain steadily throughout the interwar period; and it had begun much earlier.

After World War II, there was a dramatic reversal of the interwar trend both in productivity and in trade. But some of the long-term forces in the developing structure of world trade seem to have reappeared. In 1946 industrial output per head and manufacturing output per head were perhaps 3 per cent lower than the average of the years 1935–1938.[1] Between 1946 and 1950 both increased at the annual rate of 5 per cent and 6 per cent respectively.[2] The striking fact is not only that the rate of productivity increase appears to have been higher during the postwar years than in the interwar period,

6. At the turn of the twentieth century, 62.9 per cent of Britain's total exports were in the declining fields. Cf. Table VI.

7. Cf. Table VI.

8. *Loc. cit.*

9. Cf. A. Maizels' discussion of C. T. Saunders' paper, "Consumption of Raw Materials in the United Kingdom: 1851–1950," *Journal of the Royal Statistical Society*, Series A (general), Vol. CXV, Part III (1952), pp. 313–54, particularly p. 322; R. S. Sayers, review of A. E. Kahn, *Great Britain in the World Economy*, in *Economic History Review*, Second series, Vol. I, No. 1 (1948), p. 73.

1. L. Rostas, "Changes in the Productivity of British Industry. 1945–50," *Economic Journal*, LXII (March 1952), p. 22.

2. *Loc. cit.*

but also that virtually all industries showed improvement.[3]   Moreover, the development of productivity in British industry over the postwar period compares favorably with progress in American industry.   In the United States, output per man-hour in manufacturing in 1950 "was about 20% higher than in 1939, and the rate of increase over the post-war years was not substantially different from that in the United Kingdom."[4]   Hence, looking beyond the immediate period of reconstruction and "transition," differential rates of change in productivity over the postwar years do not appear to have been important contributing factors to Britain's long-term balance-of-payments difficulties.[5]

The volume of world trade greatly increased after World War II, and the volume of world trade in manufactures rose at an unprecedented rate.   Britain made prodigious efforts to expand her exports; yet her share of world trade from 1937 to 1950 increased by only 2.6 percentage points.[6]   Her share of world trade in manufactures rose from 19.1 per cent in 1937 to 20.2 per cent in 1951, but it fell back to 19.0 per cent in 1952,[7] despite the fact that Germany and Japan had not yet returned as strong competitors in world markets. The trend is even more apparent if we examine the trade in world manufactures of the following group of seven countries: Britain,

3. On a base of 1948 = 100, output per head in steel wire improved from 90 in 1935 to 128 in 1951.   However, output per head in iron foundries, and nonferrous metals showed rather smaller increases.   In various branches of textiles productivity either improved little (e.g., cotton and rayon weaving) or actually declined (e.g., cotton spinning and doubling).   *Loc. cit.*

4. *Loc. cit.*

5. The fact that American weekly wages in manufacturing, which were about twice the British before World War II, were about $3\frac{1}{2}$ times as high in 1951 suggests that despite any changes in relative productivity that may have occurred during the war, such changes, when translated into relative money prices, probably were not an important cause of Britain's external strains.   An index of relative United States to British export prices seems to be consistent with this view:

<div style="text-align:center">

Relative Export Prices
United States: United Kingdom
1934–1938 . . . . 0.84
1937 . . . . 0.79
1948 . . . . 1.06
</div>

Cf. G. D. A. MacDougall, "British and American Exports: A Study Suggested by the Theory of Comparative Costs, Part II," *Economic Journal*, LXII (Sept. 1952), p. 501.

6. Cf. Table VI.

7. Cf. Austin Robinson, "The Future of British Imports," *The Three Banks Review* (London), March 1953, p. 8; also given in Willard L. Thorp, *Trade, Aid, or What* (Baltimore: Johns Hopkins Press, 1954), p. 57.

Western Germany, France, Belgium, Sweden, Switzerland, and Italy. These countries conducted 50.4 per cent of world trade in manufactures in 1937, did only 46.0 per cent in 1951 and 43.6 per cent in 1952.[8] The share of the United States, on the other hand, rose from 16.6 per cent in 1937 to 28.0 per cent in 1952;[9] and her share in total world trade rose 9.5 percentage points from 1937 to 1950.[1] Canada also made substantial gains. Hence the long-term structural changes in the pattern of world trade had reappeared. But these structural changes in world trade did not operate against Britain's virtual attainment of international balance. It is not the *changing share* in world trade, but the *changing absolute volume* of a country's trade that may assist or impede its international adjustment. There is no direct relationship between a country's share in world trade and its balance-of-payments situation. France increased her share of world trade by 3.8 percentage points between 1937 and 1950;[2] she suffered from serious balance-of-payments crises. India improved her share somewhat;[3] she did not suffer from balance-of-payments crises. Belgium's share declined slightly;[4] her balance-of-payments position was strong. Regarding the longer trend, Switzerland's share in world trade declined from 1899 to 1937;[5] she did not suffer from international imbalance. Britain's share declined steadily after the 1870's; yet her external position was not weak before World War I. In so far as the mechanism of adjustment is concerned, the important thing was not that Britain's share of world trade (and in particular her share of world trade in manufactures) was declining; for it was a declining share of a rapidly *growing* trade. World trade in manufactures, for example, increased threefold between 1880 and 1913.[6] The absolute volume of Britain's trade grew rapidly and this *assisted* the operation of the international mechanism of adjustment.

Britain, France, and Germany were responsible for 60–65 per cent of world trade in manufactures in 1913.[7] Between World Wars I and II, the volume of world trade in manufactures never appreciably exceeded that of 1913; the peak of 1926–1930 was hardly above the 1913 level; the average of 1931–1935 was only 76 per cent of 1913 ;

8. *Loc. cit.*
9. *Loc. cit.*
1. Cf. Table VI.
2. Cf. Table VI.
3. *Loc. cit.*
4. *Loc. cit.*
5. Cf. Table VI.
6. *Loc. cit.*
7. Robinson, *op. cit.*, p. 57.

and the average of 1936–1938 was 92 per cent.[8] The fact that the share of Britain in particular was declining markedly within this *reduced total* was the significant factor in impeding her international adjustment.[9] The external strains of the 1930's were in considerable part the result of the need of European countries (especially Britain and France) to adapt themselves to this changing situation. It is probably for this reason (and on the implicit assumption that the volume of world trade in manufactures in the postwar era would either continue to decline or remain the same) that much emphasis has been placed upon the need of European countries to expand their *share* of world trade in order to solve their postwar balance-of-payments problems. In the case of Britain, as a result of wartime and postwar changes in assets and earning position, she was confronted with the necessity of increasing her exports of manufactures by 50 to 100 per cent if she was to balance her payments. However, Britain's share in world trade hardly changed between 1937 and 1952; yet her balance of payments by 1952 practically reached equilibrium. As a result of expansion in the volume of world trade (particularly in the volume of world trade in manufactures) Britain was able to expand her exports almost 75 per cent without the necessity of increasing her *share* in the total.[1] The changing structure of world trade therefore does not in any direct sense explain, nor is it responsible for, long-term international imbalance. It is not through changes in the share of world trade but through changes in relative cost structures, relative price levels, price structures, income levels, and exchange rates that international balance or imbalance may emerge.

IX

To sum up, differential rates of growth in productivity do not provide a satisfactory general explanation of long-term international imbalance. An hypothesis expressed in such terms, and designed to explain "the" long-term imbalance of western European countries as

8. *Loc. cit.*
9. *Loc. cit.*
1. On a base of 1937 = 100, the volume of world trade in manufactures rose to 170 in 1952. Britain's share of the total remained about the same. It is noteworthy that during this period Britain made remarkable progress in the expanding fields. Her share of world trade in this division rose 7.4 percentage points while that of the United States rose only 5.8 percentage points. Britain's share in the stable fields also rose, while her share in the declining fields fell. The United States made her greatest gains in the stable and declining industries. Calculated from H. Tyszynski, *op. cit.*, pp. 276–86.

a whole vis-à-vis the United States, appears to be refuted by the facts.
The disparity in the over-all rates of increase of productivity in manu-
facturing between the United States and the industrial countries of
western Europe was shown to be surprisingly small. The only excep-
tion was the case of Britain. Theoretically, differential rates of
change in productivity *can* cause balance-of-payments difficulties. A
more rapid rate of productivity growth in the United States, as com-
pared with Britain, may lead to a decline in United States export
prices and, given a lag in the rise of American money incomes, bring
about an increase in United States exports over imports. This might
force Britain to reduce her export prices relative to American export
prices if she is to maintain her trade in balance (or be compelled to
lose reserves). The lower the United States price and income elas-
ticities for British imports and the higher the British price and income
elasticities for United States imports (the level of money income
remaining the same, *mutatis mutandis*, in each country) the more
likely are such conditions to prevail. If such conditions were to
prevail, Britain would, in effect, tend to suffer from real as well as
monetary balance-of-payments difficulties.[2] On the basis of the
evidence, it seems probable that at certain times during the *interwar
period* Britain did so suffer. The record of the postwar years, how-
ever, suggests that differential rates of productivity growth were not
an important factor in Britain's balance-of-payments difficulties.
Under conditions of reasonably full employment and a high volume
of world trade, price and income elasticities appear to be fairly high
in the long run.[3] Under such conditions, differential rates of pro-
ductivity change are most unlikely to cause international imbalance;

2. Hicks has maintained that under such conditions the country in question
(Britain) would not suffer from real (as compared with monetary) balance-of-
payments difficulties. *Op. cit.*, pp. 122–24. This seems to me questionable on
theoretical grounds. On occasion, Hicks is prone to argue as if stability conditions
either in the domestic or in the international sphere could be used to "prove"
that a system possesses stable properties because one assumes (as he does) that
in capitalist reality the economic system does not explode. Hicks has used this
device brilliantly in *Value and Capital* and in *The Trade Cycle*, but it is particu-
larly misleading in the field of mechanism-of-adjustment theory. Schumpeter
observed that a long-run theoretical system may be explosive while the corre-
sponding reality is not; and a long-run theoretical system may be stable while
the corresponding reality is not. J. A. Schumpeter, *History of Economic Analysis*
(New York: Oxford University Press, 1952), p. 1180, n. 22.

3. For a select bibliography on materials dealing with these issues see the
present writer's, "A Note on Demand Elasticities, Income Elasticities, and
Foreign Trade Multipliers," *Nordisk Tidskrift for Teknisk Okonomi*, Oct. 1953,
pp. 39–55.

the higher the average annual rate of increase of productivity in each country, the greater the flexibility and ease of domestic adjustment to changes in international balances.

Differential rates of change of productivity in the United States as compared with the industrial countries of western Europe, expressed in terms of American improvements being concentrated on the production of products which compete most closely with those from Europe, as compared with improvements being concentrated on the production of American agricultural and other export products, also do not provide a satisfactory general explanation of long-term international imbalance, so far as mechanism-of-adjustment difficulties are concerned. While theoretically such conditions could produce monetary as well as real mechanism-of-adjustment strains, the available evidence suggests that such conditions did not prevail during the interwar period.

Nor does the changing structure of world trade provide us with a satisfactory explanation of "the" long-term international imbalance of western European industrial countries, other than the United Kingdom, vis-a-vis the United States. The evidence suggests that before World War II, changes which countries experienced in their relative shares of world trade, and the resulting balance-of-payments strains that this may have occasioned, were not so much due to the structural shifts in world demand for exports as to the ability of each country to compete in world markets for individual groups of commodities. These two forces were not necessarily related, though in the case of Britain an *indirect* relationship probably prevailed. The sheer weight of her early resource-allocation in fields which were declining in world trade must have made the problem of adjustment extremely difficult.

Differential rates of productivity growth (either quite uniformly or otherwise distributed) and the changing pattern of world trade therefore probably played an important role in the long-term imbalance of Britain vis-à-vis the United States. But her situation in many respects was quite unique. The root of the difficulty appears to have centered in Britain's relatively lower rate of productivity growth with respect to virtually all industrial countries under conditions of arrested growth in world trade of manufactures during the 1920's, and the actual decline of world trade in manufactures during the 1930's. Britain therefore encountered monetary as well as real balance-of-payments difficulties. It appears to have been erroneous to regard her external strains in the 1920's wholly in monetary terms

— the contrary position of Hicks and Keynes notwithstanding.   The overvaluation of the pound in 1925 had devastating effects precisely because it was superimposed upon a relatively weak industrial structure.   It was the painful task of long-term adjustment to the aforementioned factors which explains, in part, why Britain had difficulty in selling her goods in Canada, the United States, and nearly all industrial countries.   It seems not to have been "the increasing productivity of American industry [which] must be regarded as mainly responsible."

High average annual increases of productivity in Britain over the first decade following World War II and expansion in the volume of world trade in manufactures greatly helped to bring about virtual equilibration in her international accounts.   The experience of this and earlier periods suggests that a country's *average increase* in productivity and its *absolute volume* of world trade may be more important in assisting (or impeding) the mechanism of adjustment in equilibrating (or disequilibrating) the real variables that require adjustment than are the relative rates of change in productivity among countries or their respective share of world trade.

There appear to be no *structural* forces which have been operating abroad during the interwar period that provide a satisfactory *general* explanation of long-term imbalance of the industrial countries of western Europe vis-à-vis the United States.

<div align="right">J. M. LETICHE.</div>

UNIVERSITY OF CALIFORNIA,
BERKELEY

THE CONCEPT OF ECONOMIC SECTORS

By MARTIN WOLFE

Colin Clark's first edition of *The Conditions of Economic Progress* appeared fifteen years ago. It was greeted with some grudging admiration for the vast amount of labor it represented; it also received a shower of brickbats directed at the procedure Clark employed to show international comparisons of national income.[1] But in another section of his book, relatively unnoticed, Clark had placed a formidable tool in the hands of students of economic growth and of economic history. This was his hypothesis concerning the shift of labor during periods of economic progress from the primary to the secondary and tertiary "sectors."

By postulating the existence of "sectors," Clark implied that the economy may be divided into groups of industries so that each sector exhibits significantly different characteristics. Besides the shift of active population, these sectors were supposed to show meaningful differences in relative productivity rates and in the structure of demand for the products of these sectors during periods of rising income.

Since 1946, when I was first introduced to the *Conditions*, I have been expecting to see a brilliant outcropping of historical and analytical works using Clark's fertile concept as a point of departure. As yet they have not appeared.[2] Possibly as a result of this lack of interest, Clark has not bothered to extend and ramify this hypothesis in his second "completely rewritten" (1951) edition.

1. Critics were especially irritated by Clark's debatable methods for comparing income between industrial and pre-industrial economies. See Simon Kuznets, *Economic Change* (New York: W. W. Norton, 1953), pp. 148–53.
2. I am aware of only three exceptions which deserve notice: the works of Jean Fourastié, which will be discussed below; E. M. Ojala, *Agriculture and Economic Progress* (London: Oxford University Press, 1952); and P. T. Bauer and B. S. Yamey, "Economic Progress and Occupational Distribution," *Economic Journal*, Dec. 1951.

The scant amount of attention paid to the shifting apportionment of workers is all the more puzzling in view of the present intense interest in the field of economic growth. Clark's approach gives us a basis for comparing retarded and progressive economies in that the latter characteristically have a precipitously falling number of workers in agriculture, a sharply rising fraction in manufacturing (tending later to level out), and a steadily rising percentage in the service industries. Though many objections have been raised against the details of such an approach,[3] it does open the way for a more meaningful measurement of growth than is possible by using aggregates such as national income. Those who would place all our historical and analytical eggs in the Keynesian basket must realize that such an approach disguises many important aspects of the morphology of evolving economies.[4]

Colin Clark is not the first well-known economist, of course, to give more than a passing thought to the changing apportionment of labor. Chapter XVI ("Occupations of the People") of John A. Hobson's *The Evolution of Modern Capitalism* contains the germ of many of Clark's ideas without much sophisticated analysis, and without the notion of sectors. Even so, Hobson made good use of the meager statistical sources available at that early date (1906) to demonstrate the relative decline in agriculture (and in some aspects of manufacturing) compared with "luxury trades," public services, professions, and "other branches of non-material production"; furthermore, he used the reasoning behind what we now call "Engel's Law" to account for this phenomenon. Though some of Hobson's ideas on the subject sound rather quaint, it is apparent that Clark, had he known of Hobson's work, would not have had to reach all the way back to Sir William Petty to find acceptable parentage for his ideas.[5]

3. Contrary to Clark (and to A. G. B. Fisher), for example, a progressive economy does not necessarily go through "stages" in the sense that those workers driven or tempted out of agriculture must first find employment in manufacturing and only later, when the economy becomes still more progressive, do they move on to the tertiary. Bauer and Yamey point out that in West Africa the initial onslaught of progress brought about an immediate upsurge in distribution (tertiary) rather than in manufacturing. (*Op. cit.*, pp. 743–45.) Richard H. Holton ("Marketing Structure and Economic Development," this *Journal*, Aug. 1953) supports this point of view by reference to a similar development in Puerto Rico.

4. W. W. Rostow, for example, in *The Process of Economic Growth* (New York: W. W. Norton, 1952), discusses the size and quality of the working force (p. 91) without showing that such "quality" depends to a large extent on how labor is distributed among industries.

5. Clark does not mention Hobson, but makes much of Petty's brilliant observation "that as Trades and Curious Arts increase; so the Trade of Husbandry

"The most valuable source of information regarding the effects of modern industrialism," said Hobson, "is the study of the comparative statistics of occupation."[6] Why, then, was this tremendously suggestive hypothesis allowed to die a-borning? Apparently because Hobson, and other economists, had too many other interesting paths to explore; because the statistical information available was (and is still) too slight; and because the phenomenon involved became strikingly evident only in recent times.

Today the terms "primary, secondary, and tertiary" have become very common currency. But it is painfully obvious that those who use them have no clear idea of what they represent. Because the terms are linguistically familiar, perhaps, everybody assumes that somewhere in economic literature lies proof of their validity. But the job of working out the theory of economic sectors, as I shall try to show, has hardly begun. It is the purpose of this paper to introduce the problem of defining the sectors in the hope that others will take it up and thus help further a hypothesis which, I am convinced, will prove a great aid in clarifying some of the key issues in the fields of economic growth and economic history.

I

The circumstances which called into being the terms "primary, secondary, and tertiary" are interesting. Obviously the first two come from the standard textbook presentation of the difference between food and raw materials production, on the one hand, and "derivative" activities on the other.[7] By the 1920's "primary and secondary" were enshrined in the official terminology of Australian and New Zealand statistics.[8] Allen G. B. Fisher was the first to pin the label "tertiary" on the service industries which, in spite of their growing importance relative to the total economy, as yet had no place in the established categories of "production." Fisher coined the term "tertiary production" to dramatize the claim of this third area for consideration by those anxious to relieve unemployment, and he insisted

will decrease. . . ." Clark has labeled this concept "Petty's Law," and there does not seem to be much point in challenging the fairness of this label now.

6. *Evolution of Modern Capitalism* (London: Walter Scott, 1906), p. 383.

7. This semi-invidious comparison between economic activities which "truly produce" tangible goods and all others in a sense goes back to the physiocrats and their distinction between the "productive class" (peasants) and "the sterile and landowning classes" (everyone else).

8. A. G. B. Fisher, "Production, Primary, Secondary and Tertiary," *Economic Record*, June 1939, pp. 24–25.

that apart from transport and commerce, tertiary production could not be dismissed as "distribution," since it involved industries directly satisfying consumers' wants.[9]

Once brought into the light of day, the terms "primary, secondary, and tertiary" began to display a life of their own, even before it was established whether or not they corresponded to a meaningful tripartite division of the economy. Fisher, Clark, Fourastié, and many others began to assume in their writings that there was more to this concept than a mere group of labels.[1] The first work to examine the theoretical implications of this assumption appeared just before World War II, when Fisher marked off sectors chiefly by using the consumer's point of view and ranking the sectors in a sort of descending order of urgency. The primary sector, according to this view, is equatable with the industries producing goods most necessary for sustaining human life, and so on — with some modifications — for the other two.[2]

Though this certainly is a valid approach to the question, it leads to some annoying obscurities. And neither Clark nor Fourastié accepted Fisher's reasoning. Fourastié's entirely fresh approach is based on the relative rate of technical progress; and Clark makes no attempt whatsoever to define the sectors. In his first edition, Clark did bring up the point that primary industries are subject to decreasing returns, and secondary industries to increasing returns; but in his 1951 edition he has abandoned this distinction.[3] He continues to isolate the tertiary sometimes by viewing it as a remainder, and at other times by limiting it to those industries which produce "intangibles." Scattered about his work, however, are phrases indicating that if pressed he would admit that the terms are convenient but arbitrary.[4]

It might be well, before proceeding with a critique of Fisher's, Fourastié's, and Clark's treatments of the sectors, to indicate briefly

9. "Capital and the Growth of Knowledge," *Economic Journal*, Sept. 1933.

1. As early as 1935 Fisher was demanding "changes in the structure of production" by "constant transfers of resources" from the primary and secondary to the tertiary, even if this meant the use of government subsidies. "The Economic implications of Material Progress," *International Labour Review*, July 1935, pp. 5, 7, 17.

2. "Production, Primary, Secondary and Tertiary," *op. cit.*, pp. 31–32.

3. *The Conditions of Economic Progress* (London: Macmillan, 1940), pp. 337–42. In the 1951 edition, see his remarks on p. ix. Except where indicated, all references to the *Conditions* in this paper are to the 1951 edition.

4. As on p. 401: "For convenience in international comparisons, production may be defined as primary, secondary and tertiary."

what sort of approach I would favor. It seems to me that sectors
can be distinguished from one another by the fact that characteristic
industries in each are dominated by different factors of production,
which tend to make for a different rate of increasing productivity for
such industries. If this is true, then we could define the primary
sector as that group of industries in which an increase in productivity,
in terms of goods and services per man-hour, is limited by *natural
growth* factors, the secondary by *mechanical* factors, and the tertiary
by relatively unaided *human skill*. I am aware that this definition
seems suspiciously simple and that, perhaps, it has too symmetrical
a correspondence with the classical factors of production. But this
sort of common-sense relation to the structure of economic activity
as we imagine it is a great advantage in defining concepts which must
be widely accepted to be of value. I hope to show, furthermore, that
this approach has the advantage of consistency over previously used
definitions in that it reduces the number of hard-to-place industries
to a minimum.

I should like to emphasize the fact that this definition is based
on the firm as the unit of production — that is, on groups of firms in
industries — and thus rules out the approach to sectors through
primary, secondary, and tertiary "occupations." Unfortunately for
our purpose, it is statistics of occupations which are relatively easy
to come by; Clark and Fourastié often use them for calculations of
"sectors." But I believe it is far better to focus attention on the
firm, since technical progress works its way into the economy through
the firm. Individuals in any occupation contribute only their services;
it is the firm which combines the necessary factors and "produces."

II

Any simple definition that attempts to show how, like Gaul, all
economic activity is divided into three parts, must lead to some con-
tradictions. Subjected to an empirical test, the most meticulous and
exceedingly complex definition would remain unblemished for only a
limited time, because of the appearance of new industries, and because
many industries are changed so drastically by technical progress that
their place in the economic spectrum is not at all fixed. But con-
sistency of treatment is so lacking in Clark's work that his sectors
have more than their share of strange bedfellows.

This lack of consistency is particularly bothersome in the case of
transportation and communication, the handicraft industries ("small-

scale manufacturing"), and construction. Those who are familiar with the *Conditions* are aware that Clark places transport and communication in the tertiary sector, together with finance, the professions, government, and personal and domestic services. On the other hand, he allots the craft and construction industries to the secondary, where they find themselves associated with manufacturing, mining, and the gas and electric utilities. I should prefer to see handicrafts and construction (and especially residential construction) in the tertiary sector and the transport and communication industries in the secondary.

Clark's reason for placing transport and communication in the tertiary — the only reason I can discover — is that they are services. This brings up the question of how justified one is in establishing all "activities producing a non-material output" as one sector of the economy. To attempt a categorization of industries on the basis of whether they do or do not produce "tangible goods" may be very helpful for some purposes, but not for the purpose of marking off meaningful economic sectors.[5] In fact it is not much more useful than "sectors" of wet goods and dry goods.

There are many good reasons for associating transport and communication with the secondary sector. Their operating characteristics are quite similar to those of gas and electricity firms, which, Clark admits, are secondary industries.[6] They have enjoyed a very rapid rate of progress, much higher than that of a typical tertiary industry.[7] Their costs, estimated in terms of wages, vary greatly between progressive and retarded countries,[8] while the costs of many typical tertiary goods are surprisingly alike from country to country.[9] Finally, in recent times the fraction of the labor force in transport

5. Fisher does try, in part, to use the argument that the services are distinguishable by a separate place in the structure of demand, but this attempt, as we shall see, is not too successful.

6. *Conditions*, p. 401. That is, they have relatively high capital outlays, are large-scale enterprises, and employ mechanical processes similar to those of a factory. But on pp. 333–34 Clark refers to electricity and gas distribution in France as tertiary industries!

7. *Ibid.*, p. 325.

8. According to Clark, in the 1930's the cost of railroad transport, calculated in units of wages, could be arrayed on a scale as follows: Germany, 100; Italy, 158; China, 383; United States, 24. (*Ibid.*, pp. 322–23.)

9. In an exceedingly interesting passage Fourastié points out that in 1947 a haircut cost $1.00 in the United States and 60 francs in France, just about one hour's pay for workers in each country. On the other hand, a refrigerator cost 150 hours' pay in the United States and 1000 hours' pay in France (60,000 francs). *Le Grand Espoir du XX siècle* (Paris: Presses Universitaires, 1949), pp. 130–31.

and communication seems to be falling off — a development parallel-ing that in manufacturing — while for the tertiary industries the increase of labor "though slow, is very persistent."[1]

Another example of loose definitions (or lack of any) leading to difficulties is Clark's treatment of mining, which is "properly included in the secondary" on page 401. But why? Because national statistical offices customarily lump mining and manufacturing together? On page 314 Clark voices his qualms in a footnote: "For many purposes mining should be in the primary." Perhaps he was led to worry about mining because, like characteristic primary industries, it exhibits decreasing returns.[2] In my opinion, it is the fact that productivity in mining is linked chiefly to mechanical factors that allows us to place mining industries in the secondary sector.

Fourastié charges that Clark's division of industries into sectors is simply "enumerative" — a listing rather than an analysis.[3] And it is true that Clark makes no attempt to provide arguments for his approach, asking us only to notice that "if we give our consent . . . some remarkably convenient simplifications of analysis follow from it." But this is building on weak foundations. I must confess that it is difficult for me to understand why Clark made no attempt at a formal definition of the sectors.

I have one more sizeable bone to pick with Clark; and I trust that all this carping will not be construed as implying that I am in any sense unappreciative of the tremendously significant — and pio-neering — work he has done in this field. It is this: that many of his productivity calculations are based on prices, income, or "value added," rather than on actual goods and services. And here I know that ideal productivity calculations may not be realizeable. Obtain-ing statistics on the actual production of goods and services is fear-somely difficult. Manipulating such statistics over time brings other headaches, since the characteristics of a good (such as "an automo-bile") change so drastically over even a few years. But logic impels us to insist on the theoretical superiority of productivity defined as actual goods and services divided by hours worked. Using income, or value added, deflated by a price index to show "real output," often conceals price movements that are themselves partly functions of

1. *Conditions*, p. 403.
2. In this connection, see p. 312. On the grounds of operating character-istics, mining unquestionably belongs in the secondary. This brings to mind the use Lewis Mumford made of "The Mine" as the very symbol of the industrial revolution, the "paleotechnic era" in *Technics and Civilization*.
3. *Grand Espoir*, p. 42, n. 1.

changing productivity. "Real output" of an industry, as it is often derived, actually shows nothing more than the share of that industry in total real national income.

The danger of an approach to productivity through prices or income is increased by the fact that many tertiary firms distribute relatively large salaries, and thus may be regarded (incorrectly) as more "productive" than those in the primary sector. Misled by such calculations of "real output," some economists have urged that income in underdeveloped areas may be raised by transferring workers from "less productive" agricultural pursuits to the "more productive" tertiary.[4] The pitfalls in such an approach jump into focus as soon as we realize that changing demand has much to do with changing relative prices and income. When a nation grows rich and demands more entertainment, does this mean that the "productivity" of actors is increasing as their relative real income rises? When the relative income paid to producers of staple foods falls as national income rises, does the wheat farmer produce fewer bushels per hour of effort? What we may have here, of course, is another problem in terminology, that of deciding what is really meant by "productivity." Perhaps what is needed is Fourastié's distinction between *productivité*, based on income, and *rendement* (return), based on actual goods and services.

III

Fisher's approach, as I have indicated, is to mark off the sectors on the basis of "the structure of consumers' demand." For Fisher the primary sector consists of industries producing goods most essential for sustaining life; the secondary includes "all manufacturing activity designed to produce things for which there is a more or less standardized demand, but which could not strictly be described as essentials"; the tertiary is the remainder, which can be described also as "every new or relatively new type of consumers' demand, the production and distribution of which is made possible by improvements in technical efficiency, which releases resources hitherto required for primary and secondary production."[5] This is almost — but not quite — an identity between "primary, secondary, and tertiary" on the one hand, and "most essential, less essential, and least essential" on the other.

4. The most blatant example of such wishful thinking is in Louis H. Bean, "International Industrialization and Per Capita Income," in *Studies in Income and Wealth*, VIII, 122–23.

5. "Production, Primary, Secondary and Tertiary," *op. cit.*, pp. 31, 32.

Unfortunately for symmetry, Fisher by his own admission violates "the strict rule that there should be only one *fundamentum divisionis* in a classification."[6]   There is no problem regarding the primary sector, provided we remember that Fisher includes production *and distribution* of essential goods.   But the secondary sector, though producing moderately essential goods, is further qualified by being limited to manufacturing; and though the tertiary's "new type of consumers' demand released by improvements in technical efficiency" seems to point to luxuries and semiluxuries, Fisher insists that *all* services "which afford direct satisfaction to a consumer" be in the tertiary, including comparatively essential industries like the medical services.

The first point to be considered is whether definitions such as his (based fundamentally on utility) are applicable at all to the phenomenon of shifts in the active population.   The argument for the affirmative would run something like this: as soon as the community's income rises past the subsistence level, the proportion spent on foods declines ("Engel's Law").   If the output of agriculture remains the same while its relative income declines, the position of farmers is worsened, and they seek urban jobs.   Presumably the same forces will in time operate to lower the proportion of the labor force in factories.[7]   Therefore it seems possible, at least, to make out a case that definitions based on the changing demand schedules of a progressive economy may be an adequate frame of reference for explaining the changing proportions of the working force.[8]

This argument breaks down, however, when we ask whether there is in fact a high degree of correspondence between the sectors as Fisher defines them and the rate at which the sectors lose or absorb labor.   In Fisher's primary sector this relationship holds good only for those industries producing the basic cereal crops.   The primary,

6. *Ibid.*, p. 31.

7. But not in the tertiary industries; here the labor force may be able to expand to 70 or 80 per cent of the total, since in the utopian economy of the future the lion's share of income would go for tertiary products and since productivity in the primary and secondary will rise so steeply that demand for the products of these sectors will be met with the labor of only 20 or 30 per cent of the working population.   It is a fact, of course, that in the United States and in other advanced nations the relative numbers of the occupied population in manufacturing and mining show signs of declining already.

8. This argument is not advanced by Fisher.   It is worked out with some interesting details in Fourastié's *Grand Espoir*, chap. II; and the first propositions (the effects of rising income on the primary) are the main subject of Ojala's *Agriculture and Economic Progress*.

defined as "producing the most essential goods," does not include the same industries as that defined as "most rapidly losing labor during periods of rising income." This is obviously not the case, for example, for the products of forestry. By Fisher's definition these goods are essential and belong, therefore, in the primary. But for this category relative demand does *not* fall off as soon as income rises and therefore (productivity remaining constant) the proportion of labor in forestry will remain the same or will increase until the community becomes quite wealthy. The same argument could be applied to cattle-raising and to the production of wool and cotton, which also reside in Fisher's primary sector.

In the secondary and tertiary sectors there is the same sort of fault (for our purposes) in Fisher's definitions. The effect of rising income and productivity on the secondary, as we have seen, is that the proportion of the labor force in this sector first rises rapidly, then levels out and begins to show a slight decline. Now this is precisely the shape of the changing labor supply that we expect in the key utilities — transportation, communication, and power. But these industries are in Fisher's tertiary sector, since the secondary is restricted to "manufacturing activity" and since all services are definitely placed in the tertiary.

The basic fault in Fisher's definitions lies in the fact that we become hopelessly ensnarled if we attempt to link large numbers of goods by type of demand rather than by mode of production. This approach twists the primary sector, for example, into a very queer shape indeed, so that it contains forestry, cattle-raising, and the cultivating of basic foodstuffs, but leaves out truck-gardening, fruit growing, and dairying. His approach requires, therefore, a decision as to what degree of urgency must be allotted to the demand for each and every good and service. This would prove the toughest sort of an obstacle, since the economics of consumption is in its infancy. He who is courageous enough to undertake this task must avoid, among other difficulties, the trap of thinking that what is primary, etc., for Robinson Crusoe has the same utility for an advanced community.[9] Thus we are led to wonder whether an empirical testing of Fisher's definitions would show a fair amount of consistency only

9. Kuznets observes (*Economic Change*, p. 204) that there exists an "underlying constancy of human wants . . . over time and space . . . [men] all want food, clothing, transportation, amusement, intellectual fare, and the like. . . ." For periods of economic progress, "constancy of human wants" can only mean similarities in the changing structures of demand for different areas. But we have very little information as to whether such patterns do in fact exist.

when different arrangements of products into sectors are made for each area and each point in time.[1]

One final criticism of Fisher's definitions must be made. He places processed foods (and leather) in the primary sector, on the grounds that they enjoy a basic demand and that they have not undergone "a substantial alteration of form and character." But this leaves us the sticky problem of deciding how substantial "substantial alterations" must be before qualifying a commodity for the secondary sector.

Fisher's most recent article on this subject, unfortunately, throws cold water on the whole matter of proper definitions. Instead of making more precise the definitions of his 1939 article, he merely restates and defends them. But this is a dishearteningly lukewarm defense; Fisher now permits defining the sectors in any number of ways, though he is anxious still to make everybody see there *are* such things, and even begins by asserting that "it is a good idea to have a fairly precise notion of what we are talking about," and that "some hard core of positive meaning" is necessary for a proper classification. Towards the end of the article he blasts our hopes by admitting, with a fine show of tolerance, that "Different definitions and classifications serve different but equally useful purposes, and for each validity may properly be claimed in relation to its own purpose."[2] Like the blind men feeling the elephant, we are all to be allowed our own little say.

IV

Much more positive is the approach of Jean Fourastié to the concept of economic sectors. Fourestié's works, though of the greatest significance, are hardly known in this country.[3] To my way of thinking, Fourastié's most important contribution lies in his hypothesis of the "clash of the changing structure of production and consumption." Fourastié makes a most impressive presentation of the fact that, while technical progress and changing demand patterns both have an impact on the apportionment of labor, their effect is

1. Fisher admits this: "The production of radio sets, which was a tertiary activity in, say, 1924, should perhaps today be regarded as secondary." "Production, Primary, Secondary and Tertiary," *op. cit.*, p. 32.

2. "A Note on Tertiary Production," *Economic Journal*, Dec. 1952, pp. 820, 825, 829.

3. The most important are *Grand Espoir* (3rd ed. 1952); *La civilisation de 1960* (Collection "Que sais-je?"; Paris: Presses Universitaires, 1950); and *Machinisme et bien-être* (Paris: Editions de Minuit, 1951).

not at all identical, or even, necessarily, in the same direction. Each of these two vital forces works its effects on the economy in accordance with entirely different sets of laws. As income rises, the demand for the basic cereals, for example, falls off while, because of technical progress, the potential output of these foods continues to rise.[4] In other words, in a progressive economy consumers do not want all the offspring of goods and services which the marriage of technical progress to production is prepared to give them.[5] From this clash Fourastié traces many of the economic crises of our day — in employment, foreign trade, and prices — ills which are bound to become less critical as more and more of the population is pushed and pulled into the tertiary, where the force of technical progress is relatively slight.

Fourastié's definitions of the sectors are simple and consistent. He applies the term "tertiary" to activities that have enjoyed only slight technical progress in a given period. "On the other hand, we call 'primary' and 'secondary', activities that have enjoyed great technical progress, reserving the term primary for agricultural activities and the term secondary for nonagricultural activities."[6] Fourastié, however, does not put too much stress on his own definitions. He admits that any distinction between "great" and "slight" technical progress must be "entirely arbitrary; in fact the keyboard is continuous and we go very gradually from the top to the bottom of the scale."[7]

There can be no arguing with the logic of this definition, but unfortunately it is bound to lead to difficulties when an attempt is made to apply it empirically to the economy. In marking off the tertiary from the other two sectors, for example, the criteria must be entirely *ex post*, as Fourastié admits; we must discover the "technical progress *realized in fact* during the course of a given period."[8] In other words, this definition is not extensible to the present or the future. This must be the case, since few industries show a smooth

---

4. Incidentally, Fourastié's and Ojala's writings throw some interesting light on the long-range implications of farm subsidies.

5. Although this is kicking a dead horse, I should point out that Fourastié's work can be interpreted as providing a fresh and "dynamic" insight on the fallacy of J. B. Say's "Law of Markets."

6. *Machinisme et bien-être*, p. 124. This represents a departure from his definition in earlier works, for example *Grand Espoir*, where (p. 41; unchanged in the 1952 edition) secondary and primary were distinguished from each other in that the former showed "great" technical progress, while the latter showed only "moderate" technical progress.

7. *Machinisme et bien-être*, p. 124, n. 1.

8. *Grand Espoir*, pp. 42–43, n. 1.

curve of rising productivity on which we might extrapolate. Technical progress, after all, depends on innovation, which often presents itself in discontinuous jumps. By the same token Fourastié's approach leaves us with no place to put relatively new industries — not until we have a good idea as to how amenable they are to technical progress.

Fourastié's definitions, like Fisher's, are not crucial for the author's purposes. From the point of view of this paper they do have certain advantages, since they place main emphasis on the mode of production and since they allow for flexibility in re-allocating industries when the need arises. It should be obvious how indebted my definitions are to Fourastié's, though of course I would argue that mine are an improvement both for practical reasons and from the standpoint of theory.

<p style="text-align:center">V</p>

It remains now to point out some of the strengths and weaknesses in the logic of the definitions I have suggested. The primary, the most homogeneous sector, is defined as that group of industries in which an increase in productivity is limited by factors of natural growth. These industries all have in common the fact that no matter how efficient or intensive the use of labor, machinery, or fertilizers, a certain amount of time — over which we have little or no control — must elapse before the product is ready. Even on the largest and most progressive farms, it still takes a season to grow a crop.

This fact seems to imply that the primary sector is potentially less amenable to the force of technical progress than are the industries of the secondary sector — a notion which may jar those who are familiar with statistics, such as those of Mordecai Ezekiel, showing that *in fact* the rise in output per worker since about 1870 has been about the same in agriculture as in manufacturing.[9] But Ezekiel's figures, for our purposes, are not comparable as given, since (1) there is no proper deduction for the labor of women and children on the farm; and (2) per worker estimates, rather than per hour, disguise the fact that total annual hours worked have declined much more rapidly in urban than in rural occupations.

Another important consideration in making a "true" comparison between the primary and secondary is that we are going to have to abandon the traditional use of "manufacturing," at least as far as equating it with the secondary sector is concerned. For reasons given later on in this section, it is most important that handicrafts and

---

9. *Annals of the American Academy of Political Science*, Nov. 1936, pp. 232–33.

residential construction be removed from the secondary; and the highly progressive transportation and power industries, usually placed in the tertiary, must be credited to the secondary. Once all these corrections are granted, it seems likely that the apparent ability of technical progress in the primary to keep pace with such progress in the secondary will be revealed as an illusion.

Before more work is done on the subject, there does not seem much point in attempting an elaborate demonstration of the theoretical superiority of the secondary sector as regards technical progress. But we might be permitted to argue that if two thoroughly characteristic firms could be chosen (one primary and one secondary) to which were applied maximum "dosages" of the best machines and techniques as these became available, conceivably we might have a situation where, momentarily, the output of the firms would rise in absolutely identical proportions. But soon the man-hour productivity of the secondary firm must forge ahead, since, for the primary firm, the unit of time needed to produce a given output is tied to completion of an organic cycle. Some instances of shortening the "season" of the products of primary industry do come to mind — pigs are ready for market several weeks sooner when fed aureomycin — but for most primary activities the time limitation can be pushed back very little.[1]

In their recent and most admirable introduction to the field of economic growth, Norman S. Buchanan and Howard S. Ellis use another argument sometimes advanced against the presumed superiority of technical progress in the secondary sector: if this were truly the case, they point out, we should expect the terms of trade to turn against industrial products in favor of primary goods, while in fact we observe the reverse.[2] They use this stick (among others) to belabor the analyses of Louis Bean, H. W. Singer, and Raul Prebisch, who believe underdeveloped areas should deliberately "de-emphasize" agriculture and embark on a program of industrialization. But there are many factors besides relative productivity trends responsible for determining the terms of trade between classes of goods. It seems obvious that the higher proportional *demand* for the products of

1. An interesting attempt to distinguish between "genetic" and "extractive" primary activities was made by J. D. Black, *Introduction to Production Economics*, chap. III. "Extractive" industries, such as mining, lumbering, fishing, hunting, and "water-power utilization," are engaged in "merely harvesting" nature's gifts, while "genetic" agriculture, forestry, and fish culture "represent complete chemical recombinations" of organic materials. But this draws too fine a line for our purposes.

2. *Approaches to Economic Development* (New York: The Twentieth Century Fund, 1955), p. 262.

secondary and tertiary industries during periods of rising per capita income, for example, must offset the slower decline in the real per unit costs of *producing* primary goods. And this is especially true of foodstuffs, which suffer most in the changing "structure" of demand as income rises.

It should be said that Buchanan and Ellis are less concerned with this issue than with pointing to the fallacies in many arguments for deliberate industrialization. In addition to Bean, Singer, and Prebisch, Buchanan and Ellis might have mentioned E. M. Ojala in this connection, since Ojala, in his study *Agriculture and Economic Progress*, most specifically urges that it be a matter of world policy to transfer the "technically surplus population" from agriculture to other occupations, "and particularly to tertiary production."[3] If the "real output" of workers in the secondary sector is higher than that of those in the primary sector, then the inference (to the unwary) is plain: the smaller the ratio of workers in agriculture the better. But what would happen to such workers if consumers were not ready and able to absorb the extra goods and services being produced? To insist on industrialization in this case would be putting the cart before the horse, as Kuznets has shown: productivity and income of advanced countries are not high *because* a large fraction of the occupied population is in the secondary and tertiary; but *because a nation has a high income and productivity rate* most of its workers can be released from the necessity of producing essential foods.[4]

The secondary sector, as I suggested, can be thought of as that group of industries n which an increase in productivity is limited by mechanical factors. This "limitation" is of a totally different character than that of the primary sector, since it is constantly receding. It is the sum total of productivity-raising machines that human ingenuity has developed up to any one point in time. There is no foreseeable end to the process of mechanization until every mass-production activity is of the "push-button" variety where, perhaps, even maintenance will be carried on with a minimum of human intervention.

This approach means, in effect, an abandonment of the implied

---

3. Pp. 5, 86, and 190. Colin Clark drew similarly erroneous causal inferences in his first edition. *Conditions*, pp. 7, 11–12, 154, 179–80, and 344. However, in the 1951 edition this line of argument has been eliminated entirely, so far as I can see.

4. *Economic Change*, pp. 223–24. See also Jacob Viner's devastating comments on those who make a "dogmatic identification of agriculture with poverty." *International Trade and Economic Development* (Glencoe, Ill.: Free Press, 1952), pp. 63–71.

correspondence between "manufacturing" and the "secondary sector" which is found in Clark and in other writers on the subject. The trouble with "manufacturing" is that traditionally it means "making by hand" as contrasted with "made by God" (i.e., organic). So far as most compilers of statistics are concerned, a pair of custom-made shoes and one issuing from a shoe factory are both "manufactured." The same applies to the products of craftsmen in the Steuben Glass Company and those of workers in a highly mechanized glass bottling plant.[5] But from what has been said I think it is obvious that the handicrafts — the tool-using rather than machine-using industries — must be removed from the secondary to the tertiary sector. The mode of production of artisans has more in common with that of artists than with that of factory workers. Nobody would think of placing sculptors and painters in the secondary, even though their goods are "tangible." Indeed, in many crafts the very basis of consumers' demand is that the good produced be as far removed from typical secondary processes as possible — as any reader of Veblen knows.

Disposition of the construction industries poses an interesting problem. Until quite recently most of these activities were of the artisan type, where workers needed long apprenticeship, were tightly bound by traditional modes of operation, and used tools rather than machines. Recently, however, the use of marvelous earth-moving machinery, prefabricated units, and power saws seems to indicate a trend toward progress in the construction industries considered as a whole. Perhaps it would be possible to split this category into two, grouping some industries together as "building trades," or perhaps "residential construction," and others as "civil engineering." The latter category, including the highly mechanized pipe-laying and road-building industries, could then be placed in the secondary without a qualm. Residential construction, which few would think of complimenting (to put it mildly) by dubbing it progressive, could be left for the tertiary sector.

While some scholars might be hesitant to commit themselves concerning relative progress in the secondary and primary sectors, probably most would agree with the assertion that the tertiary sector experiences only a relatively small amount of technical progress. Unaided human skill as a factor of production is the most limited of them all. We have no reason for believing that the physical strength

5. At several points in his 1951 edition (e.g., pp. 314, 320, 366) Clark shows some misgivings about including "small manufacturing" in the secondary, but on the whole he yields to the fact that in most easily available statistics no such distinction is made.

or mental dexterity of the human animal has changed since Cro-Magnon times.

The tertiary sector includes the arts and crafts, personal and domestic service, amusement, education, and government — but *not* the utilities which, as I indicated above, are highly mechanized and should be placed in the secondary. Less sure candidates for the tertiary sector are wholesale and retail trading, finance and insurance, medical services, and the building trades. The tertiary, therefore, is the most heterogeneous sector of all, including, as Fourastié puts it, both the monk and the prostitute.

A few tertiary industries may enjoy the sort of productivity gains which Fourastié expects only in the primary or secondary. The chief reason for this is that important gains may arise through rationalizing the operations of large-scale tertiary firms, even those in no important way dependent on machinery. In retail distribution, for example, department stores and super-markets represent real progress over older types of stores. But even here we must beware of spurious progress. In super-markets, for example, some of the apparent gain in man-hour output really arises through the substitution of the customers' labor for that of the clerks.

We have seen that in the primary sector the industries are dependent on natural growth, and this would seem to mark off all firms and industries to be included in this sector in a clear and unambiguous fashion. Alas, this clarity of demarcation does not hold for the other two sectors, since, of course, hardly any firms employ either mechanical factors of production or human factors of production exclusively. It is possible to imagine firms with important contingents of various kinds of skilled labor which seem clearly secondary. Where then shall we draw the line?

Intuitively it seems easy enough to distinguish between those firms in which mechanical processes are directly and importantly linked with producing the firm's commodities or services and those firms in which machines are to a large extent ancillary.[6] An insurance agency, for example, may use typewriters, calculators, and even more complex IBM machines, but none of these is directly concerned with the provision of insurance protection. By the same token, large automobile manufacturers employ artists, clerks, economists, and hosts of others whose activities we normally associate with the ter-

6. In this case machines and "mechanical factors" include electrical devices, even though for engineers the connotations of "electronic" and "mechanical" are entirely different. This is a language difficulty hard to resolve, since we have no adjective for "machine" other than "mechanical," and it is in this sense that the word is used in this paper.

tiary, and yet nobody would doubt that the mechanical factors in such plants are crucial in determining productivity.

This sort of intuitive judgment is entirely subjective, of course, and therefore of slight use to the man who, using the Census of Manufactures, is determined to place all borderline industries in their proper sectors. Some objective criterion would be very welcome; but I am afraid that it may not be easy to achieve one which will enjoy fairly wide acceptance. I should like to suggest, by way of beginning, that those who are interested in working out empirical verifications of some concept of sectors might start with a rule-of-thumb such as the ratio between outlay for capital goods in the firm (apart from land and buildings, and averaged over a fairly short period to avoid divergent price trends) and total operating expenses. Once a large enough sample of such ratios is obtained for one industry, we might be justified in ranking them and examining the data for evidences of tendencies to cluster about one characteristic figure. After this sort of calculation is made for several industries, we might be ready to allocate the industries to either the secondary or the tertiary sector on the basis of some measure of significant differences. Such a crude procedure would be open to many objections, of course.[7] But I can only hope either that such objections will not prove insuperable or that a better basis of calculation will be suggested by somebody else. Meanwhile, such a rule-of-thumb does have the great advantage of allowing the use of easily available data, including reports to stockholders.

Deciding whether mechanical or human factors are pre-eminent at the level of the firm may allow us to decide whether some industries (as in the case of "construction") should not be split. "Industries," after all, are often only concepts which tend to vary with each statistician, while firms are likely to be entities with an objective existence. Whole industries and not firms, of course, must be the actual building blocks of our sectors, but before we are able to decide where to place these blocks, we should determine how each is put together.

VI

Any attempt to force all the country's industries into categories produces nothing more than categories, after all, and not a "true" picture of the economic corpus. But this does not mean that all definitions of economic sectors are equally useless. The phenomenon

7. For example: this device assumes that the effect on production of all types of capital goods is pretty much the same.

of the shifting labor force is a real one, we know, and there is a clear and present need for a better understanding of this development. In spite of its oversimplifications, the concept of sectors might prove a most valuable working hypothesis for economic historians no less than for students of economic growth. What a fine thing it would be if historians could compare the "structure" of the English economy before the industrial revolution with that of Russia before the first Five Year Plan! And would not the concept of the shift of labor through sectors prove useful to historians wrestling with the slippery notion of the industrial revolution itself! The insights provided by this new frame of reference would, I feel sure, make it worth our while to tackle the very obvious difficulties raised in making comparisons of "structure" across time and space.

In a new field of intellectual endeavor problems of definition often are a sizeable obstacle to progress. The recent literature concerning the theory of economic growth seems to indicate that lack of any generally accepted body of symbols may be a serious drawback. Where there is not a regrettable lack of *any* coherent framework of analysis, there is an annoying resistance to building on what has gone before. At times it seems as though the writers involved have decided that the only way to make progress is to close one's eyes to all other contributions and "go it" bravely alone.

Let my fellow historians be warned. They and others "who do not concern themselves with the abracadabra of technical economics [are wrong in thinking] that the factors making for economic growth are well-known. . . ."[8] Perhaps this floundering arises because the study of economic growth has had no Keynes. Once the "Keynesian Revolution" was launched there was no great need for further basic definitions. The frame of reference for business cycle analysis was fixed by the master, and as a result the incidence of cross-fertilization among disciples (and antagonists) was high. But in this new concern of economic theory the need for progress is so pressing that we cannot afford to wait for another intellectual messiah. A set of widely-accepted definitions, including one for the economic sectors, could be a step in the right direction.

MARTIN WOLFE.

UNIVERSITY OF PENNSYLVANIA

8. Simon Kuznets and Raymond Goldsmith, *Income and Wealth of the United States: Trends and Structure* ("Income and Wealth," Series II; Cambridge: Bowes and Bowes, 1952), p. 11.

FEDERAL CREDIT AGENCIES AND THE STRUCTURE OF MONEY MARKETS, INTEREST RATES AND THE AVAILABILITY OF CAPITAL*

By Robert Freedman, Jr.

I. Introduction

In the last thirty years significant changes have occurred in the structure of money markets in the United States. In general one can say that money markets have become more nearly uniform in various sections of the country and rate differentials between loans and securities of different maturities and risk have declined. Capital has become more widely available as well.[1]

Among the enormously complex factors responsible for the changing structure of money markets have been the federally operated credit agencies. It is with these federal credit agencies that this article is concerned. In general it is argued that the effect of federal credit agencies has been to reduce interest rate differentials between loans of different maturities or loans of different degrees of risk with the same (or different) maturity. Even where interest rates have not declined due to institutional rigidities, federal agencies have influenced the terms and conditions under which loans can be made.[2]

II. The Influence of Federal Agencies on the "Costs" of Lending

In essence, federal agencies have aided in the creation of more uniform rate patterns by tying widely dispersed and poorly related money markets into a single national money market. Federal lenders

---

1. See Table I, p. 424.

2. This article does not claim to prove the prime responsibility of agencies in rate reduction, but rather that there is good reason to believe that agency operations have played an important role in the downward movement of rates.

FUNCTIONS AND FINANCING OF FOUR FEDERAL AGENCIES
1946–1950

Agency	Major Sources of Funds in Order of Importance	Functions
1. Housing and Home Finance Agency (HHFA)		
Federal Home Loan Banks (FHLB)	Borrowings on the market.	Lends to privately owned members of the FHLB system, which lend to the public. Does not lend to individuals.
Federal Housing Administration (FHA)	Insurance premiums and other fees. U. S. Government appropriates funds for certain administrative purposes and provides capital in insurance funds.	Insures the home mortgage loans of private lending institutions.
Federal National Mortgage Association (FNMA)	(Up to 1950 borrowed from RFC). Since 1950 borrows from the U. S. Treasury. Stock now held by U. S.	Purchases and sells FHA insured and VA guaranteed and insured mortgages made by private lenders.
2. Farm Credit Administration (FCA)		
Banks for Cooperatives (Bks for Coops)	U. S. Treasury capital purchases; borrowings from the FICB; issuance of debentures to the public.	Makes loans to cooperative farm associations for operating capital, for purchase of facilities, and for commodity handling. Loans are short-, intermediate-, and long-term.
Federal Intermediate Credit Banks (FICB)	Sales of debentures to the public; U. S. Treasury owned capital stock; a federally provided revolving fund; other borrowings from private sources or the Federal Reserve.	Lends to private farm-lending organizations like production credit associations, livestock loan companies, and commercial banks. Also lends to Banks for Cooperatives. Loans are for short or intermediate periods, usually self-liquidating.

FUNCTIONS AND FINANCING OF FOUR FEDERAL AGENCIES
1946–1950   *(Cont.)*

Agency	Major Sources of Funds in Order of Importance	Functions
Federal Land Banks (FLB)	Sale of bonds on the open market, capital stock sales to borrowers, borrowings from commercial banks.	Makes direct loans to farmers. Loans are secured by National Farm Loan Associations of which the borrower is a member. Loans are for purchase of land, equipment, etc. and are long-term loans.

3. Reconstruction Finance Corporation (RFC)

	Borrowings from U. S. Treasury. Capital Stock purchases of the U. S. Treasury.	Lends at 4 per cent to public and private agencies where credit is not available at reasonable rates. Term lending to small business up to ten years is a primary function.

4. Veterans Administration (VA)

	U. S. Treasury appropriations.	Partially guarantees and insures home and farm mortgages and loans to start a business made by private lenders.

and federally sponsored or encouraged lenders have proliferated, being tied together in vast lending chains, and being underwritten, in one way or another, by the United States Treasury.[3] As a consequence lending "costs" have been reduced, making possible interest rate reductions and improved terms of lending. The remainder of this section deals with the methods by which these "cost" savings occur.

Agencies may influence the cost of funds to borrowers in two ways: by reducing the cost of loanable funds, and by reducing administrative costs.

3. For a thumbnail sketch of the function, financing and operations of the agencies being considered for the years 1946–1950, see the Chart above. The Reconstruction Finance Corporation is treated as though it were a continuing organization.

TABLE I

LOANS AND LOANS GUARANTEED AND INSURED OUTSTANDING BY AGENCY, JUNE 30, 1946–JUNE 30, 1950[1]

(*In millions of dollars*)

Agency	1950	1949	1948	1947	1946
By Direct Lending					
Farm Credit Administration	1,946	1,901	1,607	1,573	1,502
Banks for Cooperatives	242	244	230	169	162
Federal Intermediate Credit Banks	574	577	513	494	332
Federal Land Banks	930	880	864	910	1,008
Housing and Home Finance Agency					
Federal Home Loan Banks	443	363	475	239	203
Reconstruction Finance Corporation	896	833	936	947	1,534
Total	3,389	3,181	3,300	3,087	3,519
By Insurance					
Housing and Home Finance Agency	13,000	9,960	7,277	5,080	4,626
Federal Housing Administration					
By Guarantee					
Veterans Administration	4,868	3,912	3,350	2,028	394
By Mortgage Purchase					
Reconstruction Finance Corporation					
Federal National Mortgage Association	1,156	592	186	74	
Total	27,287	22,336	17,601	13,084	10,074

[1]Data largely drawn from the Annual Reports and other publications of the various agencies.

A. *Reducing the Cost of Loanable Funds*

Reducing the cost of loanable funds involves: (a) overcoming the lender's preference for cash as against securities (overcoming his liquidity preference),[4] and (b) reducing the degree of risk which the lender bears through the possibility of default.

Overcoming lender's liquidity preference. Agencies may influence the cost of loanable funds whenever they succeed in removing the lender's concern for liquidity loss. One way in which this may be accomplished is to guarantee the convertibility of securities into cash upon demand at a fixed price with nominal cost to the security holder. The Federal National Mortgage Association (FNMA) does this to the extent that it stands ready to purchase mortgages held by private lenders insured or guaranteed by the Federal Housing Administration (FHA) and the Veterans Administration (VA). The result is that private lenders may lend funds at the going rate without fear of losses due to rate increases. The Treasury bears the liquidity cost.

Although private lenders other than home mortgage lenders are not specifically aided in this way, the Reconstruction Finance Corporation (RFC) is able to lend more freely than private lenders, because this agency borrows from the Treasury at the Treasury's cost of borrowing up to a statutory limit and relends without concern for its own liquidity loss. Neither the Treasury nor the RFC is concerned with liquidity costs.

Federal Home Loan Banks (FHLB) and the Farm Credit Administration (FCA) agencies also are able to reduce rates to their borrowers even though they borrow on the market, presumably at costs reflecting the market's liquidity preference, because they lend, not as profit maximizers, but at their own operating costs without considering liquidity loss.

Increasing the money supply. Another way in which the lender's liquidity preference may be overcome and rates reduced is by changing the money supply of the economy. *Ceteris paribus*, an increase in the money supply reduces interest rates.[5] All lending agencies are

4. Assuming no default risk, lenders still may be reluctant to lend for fear that interest rates may rise, confronting them with losses because of the illiquidity of their loan. Anything that reduces fear of interest rate increases or anything which increases the availability of loanable funds in the hands of the lender reduces the premium (interest rate) which a lender must charge to overcome his liquidity preference.

5. We are dealing here only with the immediate impact of increases in the money supply. All the effects unless otherwise stated assume monetary policy will neutralize private changes in attitude toward holding cash.

financed either by borrowing from the Treasury or from the public in the open market.[6]   Although money supply effects are likely to be small and may easily be sterilized by Federal Reserve operations, they are worth mentioning.

All of the agencies which lend by borrowing from the Treasury and on the market may influence the money supply.   Funds borrowed at the Treasury (as in the case of the RFC and the FNMA) increase the money supply, whether financed out of existing Treasury cash surpluses of funds which otherwise would have remained unspent or out of Treasury borrowings from the Federal Reserve.   If the Treasury raises the money either by borrowings from the public or by special tax levies, the money supply will not be increased, but will be shifted from the public through the Treasury to the agency and back again to the public.

The FCA agencies and the FHLB, which borrow from the public, may influence the money supply under certain circumstances even though the funds are eventually returned to the public in the form of loans.   To the extent that by design or accident agencies lend in excess of borrowings on the market by drawing down their own previously acquired balances at the Federal Reserve, the money supply may be increased.   Or, if borrowing on the market exceeds current needs, the money supply is decreased if the surplus is deposited with the Federal Reserve.   These results are admittedly special cases but they are by no means unusual.

Risk pooling and shifting.   Agencies may pool risk, that is, insure themselves and their sources of funds against loss by the scope of their operations.   Or they may shift risks to the Treasury by failing to charge the premium necessary to compensate the lender for the degree of uncertainty of repayment which a given loan offers.   The first is self-insurance, the second is a subsidy.   In recent years, risk has been minimized by prosperity; thus it is impossible to differentiate clearly between risk pooling and risk shifting in analyzing the operations of an agency.

Nevertheless, the ability to pool or shift risks appears to be a significant factor responsible for the low rates charged by such agencies as the RFC, the FHLB, and the FCA agencies.   The RFC, for example, has been able to charge a uniform rate, (4 per cent) to all borrowers regardless of size or maturity of loan or geographic location.   Furthermore, federal lenders have been able to take the long view, that is, they have been able to extend maturities beyond the

6. Including borrowing from banks and other private financial institutions.

normal period acceptable to private lenders. As examples, the RFC "term" loan could be extended to ten years, and the Federal Land Banks could lend for periods up to thirty years.

Moreover, agencies are able for these reasons to interpose their superior credit between the original lender (the open market in the case of FCA agencies and the FHLB) and the ultimate borrower. This has meant that the ultimate borrower has been able to secure funds more cheaply than he would have been able to obtain them through normal banking channels.

The ability of federal agencies to reduce the risks borne by individual lenders is not limited to those agencies directly engaged in lending. The Federal Housing Administration (FHA) which insures and the Veterans Administration (VA) which guarantees private mortgage loans, by the scope of their operations and by direct federal subsidies, pool and shift risks, thus making loans available to otherwise submarginal risks. The multifaceted RFC, too, in its "deferred participation" program, does the same thing.[7]

Market stabilization. Another way in which risks may be reduced is by stabilizing the market (i.e., acting to prevent wide fluctuations in the capital value of securities). The stability of the market influences private lending because it impinges upon that nebulous psychological factor, business confidence.[8] Most important in this respect are the operations of the FHA and the VA. These agencies, in effect, in underwriting the future value of a portion of the market, give the entire market some stability. The option of private home lenders to insure future loans, too, brings with it confidence in the market although the lenders may not actually use their option. In addition to these agencies, the others which are financed by borrowing on the market, the FCA agencies and the FHLB, influence the stability of the market. A portion of the securities market, albeit a small one, may seem secure because instrumentalities of the government operate within it. Naturally, the influence of these agencies cannot be compared with that of the Federal Reserve. Also, in all likelihood, the availability of secure loans makes possible the formation of private

7. A "participation" loan is one in which the RFC and a private lender agree to share a loan. The bank earns 4 per cent on its share but has the option of selling to the RFC at will. Since the RFC buys at the will of the private lender, not only is the loan made secure, but liquidity loss need not be incurred by the private participator and the length of term need not be a hindrance to lending.

8. Market stabilization may result in more lending or lower rates because precautionary demands for cash have been reduced. If, on the other hand, transaction demands increase without an increased money supply, the net effect of pressures of both demands in opposite directions cannot be predicted.

secondary mortgage markets which, in the absence of secure loans, might not be able to operate.

B. *Reducing Administrative Costs*

Subsidies. In addition to the subsidies that compensate for risk shifting or for failure of agencies to concern themselves with liquidity loss, there are others that aid in the reduction of interest rates by decreasing the costs of operation. Nearly all of the agencies have certain administrative expenses met by appropriation. The FHLB members are provided income tax advantages over other types of banks, such as commercial banks. The FHLB as well as the Banks for Cooperatives and the Federal Intermediate Credit Banks (FICB) have federal capital which they use at no cost to themselves.[9] Agencies may invest in federal securities, which is a form of subsidy, since such income helps meet operating expenses. Most FHA insurance funds are built upon appropriated funds. All of these and many other items are evidences of subsidies which, ultimately reflecting themselves in reduced operational costs, may be passed on to borrowers.

III. Agency Influences on the Availability of Capital and Rate Patterns — the Evidence

The ability of agencies to reduce lending costs constitutes the advantage which agencies possess over private lenders. It means that agencies can lend more cheaply and/or offer better terms than private lenders. Although large private lenders may pool risks, they cannot shift the risk to the Treasury, enjoy the advantages of direct subsidies, or ignore their own liquidity costs.

The economic consequences for the money market with which we are concerned are these: (1) The cost advantage of federal lenders and private lenders aided by agencies constitutes a form of competition which has further consequences: (a) if the private lender is a "monopolist," his lending rate will tend to fall and his lending terms tend to improve; (b) whether a monopolist or not, his marginal borrowers will seek the better terms that the agency can offer. The

9. During the period considered (1946–1950) the FHLB were in the process of retiring federally held capital stock. All Federal Land Bank stock has been retired. Loss of capital stock removes the element of subsidy accruing from the interest on government securities. It does not impair the credit of this agency since, in the case of the FCA agencies, the FCA administrator can borrow up to two billion dollars from the Treasury. Thus loss of capital stock does not impair the advantage which these organizations derive from being "federal."

lender, too, will tend to restrict his loans to the better risk borrowers since he will be unable to raise his rates high enough (assuming that were his practice) to account for the probability of loss and still keep the customer. (2) Where the agency is a competitor, or is itself the only lender, the agency's cost advantage permits loan-making to marginal borrowers who would be rationed out by private lenders and to supra-marginal borrowers without adequate sources of funds.[1]

For purposes of discussion agency influences may be said either to place downward pressure on interest rate patterns or to increase the availability of capital by direct lending or by causing private enders to increase their loans.

A. Agency Aid and the Availability of Capital

The tables which follow indicate the volume of federal aid going to private borrowers. Whether the figures overstate or understate agency contribution to the availability of capital depends upon the degree to which agencies truly serve either marginal borrowers or marginal and nonmarginal borrowers located in isolated areas where lending costs are too high for private lenders. Needless to say not all borrowers are marginal since in most instances agencies lend in competition with private institutions. On the other hand, the example of profitable agency operations, in some instances, may have liberalized private lending policies and resulted in more private lending. In all events the existence of federal agency aid is assumed to be prima facie evidence that agencies have made some credit available which, in their absence, would not have been forthcoming. A discussion of the contributions of each agency to the availability of capital follows:

FCA — short and intermediate term lending — the availability of capital. Table II depicts the importance of FCA loans as compared with those made by private lending institutions.[2] Production Credit Associations (PCA) loans, to the extent that they are made in areas not otherwise served by private borrowers, increase the availability of capital. However, the PCA have a tendency to serve more densely populated, thus more profitable areas, because, although they are federally sponsored and controlled, they are privately owned. In

1. For instance, this could occur in isolated areas of such low population density that private lenders cannot exist; or where, due to institutional rigidities, the *kind* of loan required, though sound, is not made by existing lenders.
2. This excludes the very important but statistically uncalculable credit extended by machinery companies, feed, seed and cattle dealers, etc.

addition, since PCA are financed by the FICB which, although federal, are extremely conservative lenders, they are less inclined to lend to marginal borrowers whether in rural or nonrural areas. For these reasons, it is likely that the loan record tends to overstate the PCA contribution to the availability of capital. On the other hand, com-

TABLE II
SHORT AND INTERMEDIATE TERM LOANS MADE TO FARMERS BY ALL LENDERS
1920–1950
(Percentages)[1]

Date (Jan. 1)	Non-Federal	Federal[2]	FCA Agencies			Banks for Coops[5]
			Total	PCA[3]	FICB[4]	
1920	100					
1921	100					
1925	99	1	6		6	
1930	98	2	2		2	
1935	64	36	15	6	6	3
1937	55	45	19	9	4	6
1939	57	43	18	10	2	6
1941	58	42	15	10	1	4
1943	51	49	20	10	2	8
1945	52	48	24	10	2	12
1947	61	39	20	11	1	8
1948	62	38	24	11	2	11
1949	65	35	24	12	2	10
1950	65	35	24	12	2	10

[1]Computed from *Agricultural Finance Review*, Vol. 13 (Nov. 1950), p. 96. Excludes direct credit of feed, seed, cattle and machinery dealers.
[2]Includes, in addition to FCA agencies, Farmers Home Administration loans, which are intended as subsidies rather than as business type loans.
[3]Includes loans of associations in liquidation.
[4]Includes only loans to and discounts for private financial institutions.
[5]U.S.D.A., *Agricultural Statistics*, 1942, p. 702; *ibid.*, 1950, p. 672.
[6]Less than 1 per cent.

petition in nonrural areas may have liberalized the lending procedures of private banks, adding to the availability of capital from that source.

Although the contribution of Banks for Cooperatives seems to be great, the bulk of Banks for Cooperatives' funds are used as operating capital to finance inventories, payrolls, accounts receivable, supplies and the like. This type of credit is scarcely distinguishable from the commercial loans of city banks and perhaps ought not to be considered agricultural credit at all.

In summary, then, the contribution to the availability of short-term capital by agencies is seen to be relatively small, despite the presentation in Table II, when it is realized that: (1) adequate figures for dealer credit, at present unavailable, would sharply reduce the proportion of federal credit shown in Table II; (2) PCA credit tends to be competitive to, rather than supplementary to, private capital; and (3) most Banks for Cooperatives' loans are probably better classified as commercial rather than agricultural credit.

TABLE III

FARM-MORTGAGE DEBT: PERCENTAGE HELD BY PRINCIPAL LENDERS, SELECTED PERIODS, 1910–1950[1]

Year	Non-Federal Lenders	FLB's	Other Federal Lenders
1910	100		
1920	95	1	4
1930	81	7	12
1935	62	13	25
1937	56	14	30
1939	56	11	33
1941	57	12	31
1943	59	12	29
1945	65	11	24
1946	68	9	23
1947	73	7	20
1948	76	6	18
1949	78	5	17
1950	79	4	17

[1]Adapted from *Agricultural Finance Review*, Vol. 13 (Nov. 1950), p. 87.

FCA aid to long-term lending — the availability of capital. Table III indicates the importance of the FLB in the long-term agricultural credit field.[3] FLB not only lends in areas poorly served by private lenders, but also tends to make the more marginal loans. Small commercial banks operating in rural areas are likely to know their customers and choose the least risky. They are also in a posi-

3. The high percentage of FLB loans in the 1930's is due to the decline in private credit during the depression and the assumption of certain risks by the FLB in that period. The recent decline in FLB loans is due to high repayment rates plus a very conservative lending policy. FLB lends up to 65 per cent of the long-term value of the property to be purchased.

tion to watch the farmer's progress and supervise the loan when necessary. The FLB, located in only twelve cities, cannot be in as close contact with the customer and may neither be able to select as good a loan initially nor prevent its deterioration by control later. In addition, it is probable that the better-risk borrower would approach a commercial bank before a National Farm Loan Association,[4] since the low-risk farmer is likely to have established a line of credit with the local bank for other purposes. Since commercial bank lending policies have not been so conservative as FLB policies in recent year,[5] borrowers whose credit rating permits better terms from private lenders would prefer to deal with them.

For these reasons it is likely that FLB loans are for the most part contributions to the availability of capital although, no doubt, there are areas of direct competition with private lenders where FLB aid tends to replace rather than to complement private credit.

FHA, VA and FNMA — the availability of capital. The FHA insures and the VA guarantees privately written mortgages against losses while the FNMA is a federal secondary mortgage market to which FHA-insured and VA-guaranteed documents can be sold under certain circumstances.

These agencies may increase the availability of capital in two ways: (1) Since FHA and VA terms, down payments, length of amortization and interest rates, usually require smaller monthly charges than conventional mortgage loans, borrowers, otherwise unable to afford mortgages, may be enabled to borrow.[6] (2) Since the VA and FHA guarantee the principal of the loan against loss, borrowers, previously considered too risky, may be extended credit.

However, guarantee against loss of principal is not enough in itself to insure liberal home lending. Low rates of interest which attract borrowers repel lenders, unless the rates are no lower than those on alternative but comparable loans.[7] Further, in the case of the VA, where guarantees are on only 50 per cent of the mortgage,

---

4. All FLB loans are made through the NFLA which, in effect, guarantees the loan made by the FLB. There are over 1200 such associations. The borrower makes loan application there rather than directly with the FLB.

5. See n. 3, p. 431.

6. It is true that longer amortization and smaller down payments mean greater total cost to the borrower for a given rate of interest. But many potential home owners are primarily concerned with the monthly cost of housing.

7. That is to say, rates on risk-free government securities of comparable maturity can differ from mortgage rates only by risk premiums and administrative costs.

the unsecured portion represents a lending hazard.[8] In addition, long amortization periods, important inducements to borrowers, are of no interest to lenders wishing to remain liquid.

The FNMA offers one solution to these latter problems. That agency, by purchasing the insured or guaranteed loan at par, allays the fear of loss on long-term loans which might occur as a consequence of increase in the interest rate, and shifts the risk on the unsecured portion of VA loans to the Treasury. Thus any consideration of the VA and the FHA in terms of the influence on the availability of capital must include the FNMA.

Table IV indicates the importance of the FHA in making capital available. It can be assumed that by and large FHA insured mortgages are made to marginal borrowers because they are less flexible, they involve a certain amount of red tape and they do not leave as much discretion in the hands of the lender. Thus when they are made, they probably add to the availability of capital.

VA loan data found in the same table probably exaggerate the contribution of that agency to the availability of capital. Not all veterans are marginal borrowers, although many, no doubt, are. However, veterans gain considerable financial advantage by taking a VA loan while lenders suffer no financial disadvantage. In the postwar period VA maximum rates were 4 per cent as against FHA rates of 4½ per cent. The difference in rates can be considered the "insurance premium" of one-half per cent paid by the VA to itself. FHA insurance premiums are paid by the borrower. The lender is no worse off since in neither case does he retain the "insurance premium." In addition, there are other benefits to the borrower which make a VA loan advantageous to him while the lender, being essentially sympathetic with veterans and suffering no financial disability, tends to make VA loans to both marginal and nonmarginal borrowers.

Table IV also clarifies the role of the FNMA. The postwar period first showed declining rather than rising rates of interest. FNMA purchases of FHA-insured mortgages moved in the same direction, declining when rates declined, increasing when rates rose.[9] After VA loans became eligible for purchase in July of 1948, FNMA

8. VA guarantees often are combined with FHA insurance on a given loan which effectually covers the entire loan.
9. This is overly simplified since FNMA purchase authority and other institutional factors played an important role in determining FNMA purchases.

purchases skyrocketed in 1949 and 1950.[1] It is noteworthy that in the years of spectacular FNMA purchases of VA's loans, the percentage of VA's loans written dropped sharply. The most important reason is to be found in rising interest rates which made 4 per cent VA loans quite unprofitable.

TABLE IV

RELATIVE IMPORTANCE OF FHLB MEMBERS, FHA, VA AND FNMA
IN MORTGAGE MARKET, 1939–1949; RATES ON LONG TERM
U. S. GOVERNMENT SECURITIES, 1939–1949

Year	All Mortgages Written[1] (millions)	Percentage of All Mortgages Written			FNMA Holdings[2] (millions)	Rates on U. S. Gov't Securities[3]
		FHLB	FHA Insurance	VA Guarantees		
1939	$3,506.6	23	26		$125.2	2.36
1940	4,031.4	25	25		163.0	2.21
1941	4,732.0	25	25		194.1	1.95
1942	3,942.6	23	29		215.5	1.93
1943	3,861.4	27	24		72.2	1.96
1944	4,605.9	29	19		59.5	1.94
1945	5,649.8	31	12	1	10.0	1.60
1946	10,589.2	31	7	10	6.4	1.45
1947	11,728.7	29	10	13	4.9	1.58
1948	11,882.1	28	28	8	51.0	2.00
1949	11,818.0	28	32	6	464.5	1.71

[1]U. S. Department of Commerce, *Survey of Current Business*, *July 1950*, p. 34. Data prior to 1939 not available.
[2]Housing and Home Finance Agency, *3rd Annual Report*, p. 74. Until July 1948 only FHA loans were purchased.
[3]Board of Governors, Federal Reserve System, *Banking and Monetary Statistics*, p. 468; *Federal Reserve Bulletin*, Nov. 1945, p.1138, Jan. 1948, p. 75, Jan. 1951, p. 66. All figures from 1942 are 7–9 year taxable issues.

The FHLB system — the availability of capital. The Federal Home Loan Banks supervise and lend to members of the FHLB system, largely, but not exclusively, Federal Savings and Loan Associations. The FSLA were established as a means of channeling savings directly into the mortgage market, since mutual savings banks, the only specialized institutions dealing in mortgage credit, were largely confined to New England.[2]

How successful this system has been in making capital available

1. Figures for the first half of fiscal 1950 show over $800 million in FNMA holdings, mostly VA loans. HHFA, *Third Annual Report*, p. 74.
2. Private individuals, commercial banks and insurance companies have been the traditional sources of mortgage credit in the United States.

beyond that which would have come from traditional sources is difficult to estimate. Table IV indicates the volume of business done by these specialized institutions.

FHLB members have more than held their own in the growth of the mortgage market which means either that they have "captured" business which would have gone to other lenders or that FSLA have encouraged more lending and borrowing than would otherwise have been the case, that is, increased the availability of capital.

Evidence to support the view that more capital has been made available in the postwar period is very flimsy for these reasons: (1) Since 1935 over two hundred new institutions have joined the FHLB system.[3] This could, in itself, account for the increase in loans. (2) Although the FHLB are a source of relatively inexpensive and stable credit, the banks have not been a very important source of funds in recent years.[4] Thus members compete for savings as the prime source of loanable funds (although they do borrow from commercial banks as well as from FHLB) and are unable to offer loans at sufficiently lower cost to attract borrowers who would otherwise be unable to finance a mortgage. (3) Even the up-to-date lending procedures, lack of fear of being in debt, tax exemptions and publicity provided by the FHLB Board do not mean that the FHLB members have done more than capture business otherwise done by private persons and institutions. It is true, however, that the formation of FSLA in areas where no specialized institutions for mortgage credit exist has probably helped develop a "more perfect" mortgage market. In this sense, borrowers who have refused to do business with "monopolies" because of unfavorable terms or unpredictable behavior, may have been encouraged to borrow. One may conclude that the FHLB system probably has not increased the availability of capital greatly.

The RFC — the availability of capital. The RFC is the only federal agency concerned with aid to business enterprise. The fundamental problem of whether RFC loans represent a net increase in the availability of capital depends upon how well that agency conformed to its rule to lend only where capital was unavailable at reasonable rates.

This policy criterion suffers from lack of clarity. What are reasonable rates? In practice borrowers would need only to demon-

---

3. *Housing Statistics Handbook*, p. 123. Although many that joined are old institutions, others were newly created.
4. *Ibid*, p. 125. Advances outstanding have ranged from $100 to $300 million while loans have been from $1 to 2 billion.

strate that local banks would not lend for the necessary term at reasonable rates in order to qualify for an RFC loan. Borrowers were not required to canvass lenders in nonlocal areas, nor expected to utilize equity financing, nor required to draw upon the personal resources of owners. Whether forcing borrowers to incur the "costs" of complying with more stringent conditions of eligibility would have defeated the purposes of the RFC cannot be determined in any single case, but the difficulty of being entirely certain that RFC loans represented a net addition to the availability of capital is clear.

TABLE V

COMPARISON OF AMOUNT OF BUSINESS LOANS BY COMMERCIAL BANKS AND BY THE RFC, 1939–1947[1]

(In thousands of dollars)

	Commercial and industrial loans by insured commercial banks		Loans to business enterprises by RFC		Grand total
	Total	Percent guaranteed by RFC participations	Total	Percent of grand total	
June 30, 1939	$5,511	1.4	$121	2.1	$5,631
June 29, 1940	6,360	3.1	135	2.1	6,492
June 30, 1941	8,205	2.8	180	2.1	8,382
June 30, 1942	8,868	. . .	259	2.8	9,127
June 30, 1943	6,888	2.0	435	5.9	7,321
June 30, 1944	7,407	.9	498	6.3	7,904
June 30, 1945	7,502	.9	402	5.1	7,903
June 29, 1946	10,336	1.5	301	2.8	10,635
June 30, 1947	14,768	9.4	291	1.9	15,050

[1]U. S. Congress, Senate, *Hearings before a Special Subcommittee of the Committee on Banking and Currency on Senate Resolution 132*, U. S. Senate, 80th Congress, 2d Session, January 14–25, 1948, p. 580.

Table V shows the relative importance of the RFC as a business lender from 1939–1947. Only in the one year, 1944, did the RFC's share of business aid exceed 6 per cent of the total. This relatively large volume was due to the national defense aspects of the loans made in that year. Ordinarily from 2 to 3 per cent of business long-term loans are RFC loans. In 1947, the RFC participated with private banks very heavily. This is attributable to the high return on risk which a bank receives and is dependent on low market rates in the economy.[5]

5. Since the bank can shift its portion of the loan to the RFC at will, the loan is risk free and need only be the best possible use for money.

Table VI indicates that the RFC made a much more important contribution to term lending to small business in 1941 than to any other group of borrowers. While the percentage of the total amount loaned to this group was small, 91.9 per cent of all RFC loans went to small business. The relative neglect by commercial banks of term loans to small business is attested to by the small percentage of the total number of loans and the small percentage of the total amount loaned that went to that group.

TABLE VI

PERCENTAGE DISTRIBUTION, BY SIZE OF LOAN, OF THE NUMBER AND AMOUNT, OF TERM LOANS MADE BY COMMERCIAL BANKS AND BY THE RFC, 1941[1]

| Size of Loan | Percent of Total Number | | Percent of Total Amount | |
	Commercial Banks	RFC	Commercial Banks	RFC
Under $5,000		37.6		1.9
5,000–10,000	9.0	15.9	.1	2.7
10,000–25,000		19.0		7.3
25,000–50,000	16.7	11.2	.7	9.4
50,000–100,000	10.4	8.1	1.1	13.4
Sub-total	36.1	91.8	1.9	34.7
100,000–500,000	33.0	7.2	12.0	33.3
500,000–1,000,000	12.7	.6	12.3	10.1
Over 1,000,000	18.2	.4	73.8	21.9
Sub-total	63.9	8.2	98.1	65.3
Total	100.	100.	100.	100.

[1]Source same as for Table V, pp. 581–82.

It is not surprising that most RFC money should have gone to large business, but the surprising thing is that although 37 per cent of the loans made were for less than $5,000, as much as 34.7 per cent of the total amount disbursed went to small business.

There is every likelihood, given the size of most RFC loans and the conservative lending policies of banks, that the bulk of RFC loans do represent a net addition to the availability of capital.

B. *Agency Influences on Patterns of Interest Rates*

It is an hypothesis of this paper that the "cost" advantage of federal agencies has not only increased the availability of capital but

has aided in reducing the patterns of interest rates charged by private lenders.

However, there are at least two other distinguishable forces which may bring the same result. One is federal monetary policy which influences the yields on risk-free federal securities; another is economic development which bears upon the supply and demand of private credit.

The first poses no special problem and can easily be isolated. Monetary policy cannot influence regional rate differentials, as can agency operations, because it deals with national markets alone. Further, should interest rates in any locale change by more than the change in security yields of relevant maturities, it cannot be due to monetary policy. Monetary policy cannot influence (directly) the risk premiums added to the basic cost of money, which are the same for all uses of money for a given period of time. Declines in yields beyond those justified by federal monetary action must be due to changes in some component of lending cost other than the cost of foregoing liquidity, such as risk or administrative costs.

Economic development, however, poses a nearly insoluble problem to the investigator attempting to isolate the factors responsible for interest-rate declines. Economic development tends to narrow regional rate differentials as do agency activities. Economic development results in increased flows of both private and public capital, alleviating the capital shortage that is probably partly responsible for regional differentials. Also, it may result in the development of branch banking or the establishment of new financial institutions tending to break the hold of credit monopolies.

Economic development also can reduce over-all lending rates by helping to reduce default risks. Prosperity, which has been a feature of the rapid postwar economic development, has reduced risks and has been partially responsible for interest-rate declines.

Because of the overwhelming importance of economic development in rate reductions, it is only by inference that one can evaluate the role of federal agencies. However, there are these reasons for believing that agencies have played their part:

(1) Agencies usually charge the same rate to all borrowers (for a given type loan) regardless of risk or administrative cost. Where agency rates are below private rates, better-risk private rates must be reduced to meet agency competition and poorer-risk borrowers, being denied credit, must seek credit with federal agencies.

(2) Agencies usually (not always) charge the same rates in differ-

ent regions. For the same reasons, this will cause regional rate differentials to decline.

It is argued below that agencies are to some degree responsible for interest rate declines.

FCA short and long term lending — interest rate patterns. There is no study of short-term agricultural lending rates to buttress or refute the contention that FCA agencies have helped reduce private short-term rates. However there is no doubt that private lenders consider the PCA to be competitors and adjust their own lending rates to meet the competition.[6]

It is unlikely, however, that the PCA could force private rates down very greatly except where "monopoly" bankers are exploiting their clients. Short-term rates are of necessity high because of the expenses connected with administering loans coming up for renewal at short intervals.

So far as long-term credit is concerned, not only have farm mortgage rates fallen, but regional rate differences have narrowed greatly in the last forty years.[7] There is little doubt that much of the credit for the narrowing of rates is due to economic development and that credit for the over-all decline in rates is due to falling yields in United States Government securities.

However there is some reason to feel that the FLB have played their part in rate reduction. The FLB rate is below the average of private rates everywhere.[8] Even though this does not prove that the FLB have pulled private rates down, it infers that FLB rates act as an anchor preventing their rise. After all, even though private lenders do not ordinarily consider FLB to be competitors, FLB rates stand as a warning to private lenders that borrowers do, in the end, have alternative sources of credit.

The FHA, VA and FNMA — interest rate patterns. The FHA can certainly be credited with reducing regional rate patterns on home mortgages. Table VII shows the change in mortgage rates in nine regions in the United States in 1931 and 1936. Rates declined in all nine regions, and differentials between regions were also drastically reduced, as shown in the table. In 1934 the FHA introduced

---

6. Interviews with several country bankers in the central New York area substantiate this belief.

7. *Agricultural Statistics*, 1942, p. 697; *ibid*, 1950, p. 662. Figures from 1900–1950 for nine regions and for the United States as a whole illustrate this fact.

8. Between 1946 and 1950 the FLB rate everywhere was 4 per cent while the average private rate on long-term farm mortgates for the United States was 4.6 per cent with no region below 4.3 per cent.

TABLE VII

Home Mortgage Interest Rates in Selected States, 1931 and 1936[1]

| States | Effective Average | |
	1931	1936
New England		
Connecticut.....................	6.2	5.9
Vermont......................	6.8	6.0
Middle Atlantic		
New York......................	6.4	5.7
Pennsylvania..................	7.8	6.1
East North Central		
Michigan......................	7.1	6.2
Indiana.......................	7.8	6.4
West North Central		
Iowa..........................	7.1	6.2
North Dakota..................	9.8	6.9
South Atlantic		
North Carolina.................	6.5	6.0
West Virginia..................	10.2	6.2
East South Central		
Kentucky......................	8.4	6.2
Tennessee.....................	15.3	6.3
West South Central		
Louisiana......................	7.5	6.0
Texas.........................	10.5	7.2
Mountain		
Wyoming......................	8.2	6.8
New Mexico...................	11.6	6.6
Pacific		
California.....................	8.8	6.7
Oregon........................	10.3	6.6

[1]Source: Unpublished dissertation by Earl S. Garver, "Mortgage Insurance," deposited in the Yale University library, 1943, p. 669. The highest and lowest effective rates charged in each area in 1931 are shown.

mortgage insurance everywhere at 5 per cent, and that rate was retained through 1936. Earl S. Garver[9] credits the FHA with the results found in Table VII but feels that credit is also due the Federal Home Loan Bank System and the RFC Mortgage Company (the then operating federal secondary mortgage market). The FHLB system was a source of funds in areas traditionally short of private savings and the RFC Mortgage Company made insured loans liquid — in effect made the legal FHA rate a realistic one in overcoming liquidity preference.

Whether the VA has played a part in the pattern of mortgage rate reduction cannot be answered with any certainty. For one thing, the VA came on the scene in 1945 long after the FHA, the FHLB and the FNMA had done the essential work of narrowing rate differentials between regions. Thus there was little for the VA to do. In the second place, so far as interest rates are concerned, the VA and FHA differ only in that the FHA borrower pays a one-half per cent insurance premium. The VA could not have forced rates down on conventional (uninsured) loans beyond that already accomplished by the FHA, since risks on conventional loans must be paid for by an interest premium.

On the other hand, the analogy with the FHA ends here. Many borrowers who are in no sense marginal are eligible for a VA loan. The best of these borrowers would get the lowest conventional rates offered by a lending institution, which could be (under certain circumstances) virtually the same as VA rates. If, owing to monopoly, (much less likely since the advent of the FHLB) the rate offered is substantially higher than the VA rate, the borrower may insist upon VA terms. Although a lender cannot be forced to make a VA loan, he might be forced by fear of bad publicity either to make a VA loan or a comparable conventional one.

The FNMA, like the RFC Mortgage Company which preceded it, plays its part in the maintenance of low rates. Being willing to purchase FHA and VA loans at par, that agency makes it unnecessary for the private lender to concern himself with liquidity loss. Thus, given low rates on risk free federal securities, the FHA and VA loan can be made at low rates.

Conventional mortgages, however, cannot be sold to the FNMA.

9. Earl S. Garver, "Mortgage Insurance," unpublished Ph.D. dissertation, Yale University, 1943, p. 671. Garver suggests that FHA insurance also encouraged the development of private secondary mortgage markets which increased the liquidity of mortgages and made the task of the RFC Mortgage Company easier.

It is only because of the development of private secondary mortgage markets (insurance companies primarily) that long-term low-interest-rate conventional mortgages can be made.

The FNMA and the FHA are also responsible for the low rates on conventional mortgages because they are partly responsible for the existence of private secondary mortgage markets. Private secondary mortgage markets are possible because of the security and liquidity provided by FHA mortgages eligible for sale to the FNMA. These markets need bear no default risk nor be as much concerned with liquidity loss as is a bank.[1] Since most of the portfolios of banks are secure and probably liquid, they may hold conventional mortgages. Conventional mortgages are likely to bear little risk, being made for supra-marginal borrowers, and are likely to be relatively liquid since, given a widespread secondary market, liquidity and shiftability are identical from the point of view of a particular holder.

The RFC — interest rate patterns. Table VIII indicates the role of the RFC with respect to the patterns of rates on long-term business loans.

TABLE VIII

INTEREST RATES BY SIZE OF LOAN FOR TERM LOANS
BY COMMERCIAL BANKS, NOVEMBER 20, 1946[1]

Size of Loan	Term Loans
Less than $500	8.9
500–999	7.7
1,000–4,999	5.9
5,000–9,999	5.1
10,000–24,999	4.6
25,000–49,999	4.2
50,000–99,999	3.8
500,000–999,999	2.4
1,000,000 and over	2.8
Grand Average	2.8

[1]U. S. Department of Commerce, *Survey of Current Business*, July 1950, pp. 581–82.

1. Even without the ultimate liquidity of the FNMA, insurance companies can hold some mortgages so long as they are secure. Insurance companies are not so much concerned with liquidity as banks because their liabilities are of much longer term than demand deposits. Insurance companies, because of the distribution of maturities and the predicatable nature of liabilities, may hold mortgages to maturity and need suffer no capital loss. A bank might require liquid assets quickly and might have to sell at a loss.

As late as 1946 the smallest "term" loans in the country as a whole carried rates higher than 8 per cent, while the bulk of RFC loans made for less than $5,000 dollars carried a uniform rate of 4 per cent. The RFC might have had some influence in reducing business rates charged by private lenders but this is difficult to prove.

In addition to business loans, the RFC has intervened in the long-term bond market. In recent years the RFC has purchased revenue bonds issued by the Pennsylvania Turnpike, the Queens Midtown Tunnel in New York, the Brooklyn-Battery Tunnel in New York and others. By 1948 the turnpike bonds had been sold at a profit.[2] To have achieved this result, the RFC must have bid in at low prices, although higher than those of competitors, and resold at still higher prices with the result of driving bond yields down. Of course, the RFC cannot enter into competitive bidding everywhere, but it probably has had some influence in raising the prices of public authority bonds and driving bond yields down where it can bid. Again whether the RFC policy has had a permanent impact upon private rates cannot be proved.

Further, Robert V. Rosa claims the RFC has diverted funds from cheap to dear money areas.[3] In other words it would seem that the RFC's uniform interest rate policy has played a part in the reduction of interest rate differentials between geographic areas, although there is no proof for this assertion.

It may be true that the RFC has helped bring about uniform interest rates, but given the relatively unimportant volume of lending done by the RFC, the influence of that agency can easily be exaggerated.

IV. Conclusions

Federal credit agencies have played an important role in the development of a national money market both by providing their own capital and by influencing the rates, terms, and possibly the availability of wholly private credit. In more recent years, economic and commercial development has tended to obscure the importance of federal agencies by aiding the growth of branch and chain banking and contributing to the emergence of new financial institutions in areas hitherto dependent on federal credit.

2. *Hearings before a Special Subcommittee of the Committee on Banking and Currency, on Senate Resolution 132*, U. S. Senate, 80th Congress, 1st Session, pp. 431–32.

3. Robert V. Rosa, "The Monetary Power of Some Federal Agencies Outside the Federal Reserve System," unpublished Ph.D. Dissertation, University of Michigan, 1941, p. 296.

However, although the money market has become more truly national in scope, it is by no means certain that federal agencies have become obsolete or that continued subsidy to borrowers is unnecessary and undesirable. It is particularly true that rural areas tend to be underserved partly because the size of the market does not permit wide risk pooling and partly because it is by no means certain that country bankers are capable of or interested in maximizing profits. It is also true that where long-term lending is concerned, the problems of liquidity facing private lenders, rural and urban, still prevent an entirely adequate flow of credit.

The only completely general conclusion which can be drawn is that agencies have played and still play an important role in the maintenance of low rates and availability of capital. The degree of importance may be disputed. It is hoped that this analysis of the *methods* by which agencies provide credit and the tentative *appraisal* of their success will elicit a more thorough investigation of credit agencies. A reappraisal of agency financing techniques and lending policies might well lead to a better "fit" between agency objectives and performance.

                                        Robert Freedman, Jr.

Colgate University

A NOTE ON SOVIET CAPITAL CONTROVERSY*

By Alfred Zauberman

An observer of developments in Soviet economics must register an important event. After a prolonged reticence the guardians of Soviet economic doctrine have made a pronouncement in the great Debate — probably the only one which has been raging with singular heat for years in Soviet theory. Although it has been made anonymously, no mistake can be made as to the weight of authority which it carries.[1]

The point at issue is: what criteria are there, if any, in the Soviet economic model for selection of investment alternatives? Of necessity this reduces itself to a more fundamental question, viz., what are the methods and standards of measuring the efficiency of capital in a highly collectivized nonmarket system?

Engels insisted in a famous passage of the Anti-Duehring that in the postcapitalist community people would be able to manage everything without the intervention of value. As early as 1844 he assured his readers that the "balancing out of useful effects and expenditure of labor would be all that would be left in the Communist society of the concept of value, as it appears in political economy."[2] The innocence of Engels' proposition must be abundantly clear today

1. *Voprosy Ekonomiki*, No. 3, 1954. English translation in *Soviet Studies* (Glasgow), VI (Oct. 1954), 201–17. The authorship of the Statement has not been disclosed, presumably to give it the imprint of official anonymity. But Professor Notkin's new book which came into my hands when this Note was going to print (*Voprosy Opredieleniya Ekonomicheskoy Effektivnosti Kapitalnykh Vlozheniy v Promyshliennost SSSR*, Moscow, 1953) helps to trace the paternity of some of its most important parts. By comparing the text one can detect without great difficulty Notkin's hand in formulating the arguments for rejection of those schools of thought which were ready to accept some uniform rate of capital efficiency throughout the economy or its individual branches (pp. 89 seq.). Notkin has also clearly helped in establishing one of the few positive principles, i.e., that of direct commensurability of current cost and capital outlay on Marxian grounds (p. 93). On the other hand, it seems that it was also he who was instrumental in formulating those passages which have left the vexed problem in its drift.

2. Deutsch-Franzoesische Jahrbuecher, p. 95, as quoted in the *Anti-Duehring*, Calcutta edition (1943), p. 294.

to his Soviet pupils. In a system where no market objectifies for the entrepreneur the price of capital, establishing the economic rationale of investment choices has proved a particularly tough nut to crack. Even the project-maker, i.e., the one whose job it is to make the most sensible choice among alternative ways of reaching the prescribed output, within the very restricted field left to him by the sovereign Planner, must feel uncomfortable groping in uncertainty. It was he who challenged the theoretician to help him out of his predicament. In fact that is how the great Debate began.

That Debate has attracted more attention in the West than any other happening in Soviet economic thinking. We may be absolved, therefore, I believe from retelling its story in greater detail.[3]

The pronouncement admits that as far as basic concepts at issue in the Debate are concerned, the dychotomy into planning and project-making is irrelevant. Those who try to play safe by referring the project-maker to the Plan for guidance have been given to understand that they were evading the problem. Planning capital construction itself — the pronouncement teaches — "calls for assessment of capital efficiency of investment alternatives within the framework of the national economy as well as of the individual enterprise" (pp. 100–1). After all, unless those in authority would be prepared openly to grant the Planner exemption from economic criteria he, no less than his project-makers, has to rely on essentially the same set of basic principles of choice.

Thus the pronouncement has widened the scope and consequently also the importance of the Debate. But involving in its argument the sovereign Planner himself was bound to add to embarrassment and difficulties in the search for the solution.

Nobody would expect that when the final moment of resolving

3. Cf. in particular: H. Hunter, "The Planning of Investments in the Soviet Union," *Review of Economics and Statistics*, XXXI (Feb. 1949), 54–62. J. M., "Some Recent Developments in Soviet Economic Thought: Economic Choice Between Technological Alternatives," *Soviet Studies*, I, (Oct. 1949), 119–27. A. Zauberman, "Economic Thought in the Soviet Union: I. Economic Law and the Theory of Value," *Review of Economic Studies*, XVI, 7–10; idem, "The Prospects for Soviet Investigations into Capital Efficiency," *Soviet Studies*, I (Apr. 1950), 328–33. Charles Bettelheim, "The Discussion on the Problem of Alternative Investment Projects," *ibid.*, II (July 1950), 22–42. Maurice Dobb, "A Note on the Discussion of the Problem of Choice Between Alternative Investment Projects," *ibid.*, II (Jan. 1951), 289–95. Norman Kaplan, "Investment Alternatives in Soviet Economic Theory," *Journal of Political Economy*, LX (Apr. 1952), 133–44. Gregory Grossman, "Scarce Capital and Soviet Doctrine," this *Journal*, LXVII (Aug. 1953), 311–43.

the argument arrived, the impotence of the model for formulating quantitative criteria of choice would be frankly admitted. Nor was it. Platitudinous pseudo-criteria suggested by some writers — such as "strengthening and expanding the superiority of the forces of socialism over those of capitalism" or "the speediest construction of a communist society" — have been declared "qualitative" only and hence insufficient, much as they are in vogue in Soviet writing. Yet we shall see in a while how timid the pronouncement is in its search for *quantitative* criteria.

To begin with, "the view according to which the law of value, the profitability of the individual plant was a regulator of allocation of investment resources has been found wrong" (p. 101). This finding has even gained the distinction of praise as being the "positive result of the Debate" (*ibid*). Since, however, it appears that the law of value is powerless with regard to the Planner's decisions, not very much could be devised in the way of that quantitative guidance which seemed to be promised earlier by the pronouncement. Instead, the Planner is referred to a collection of sundry desiderata.

Now, what about the project-maker? He is to be led by considerations of labor-saving and profitability, as between alternative project variants, *if* they are "more or less equivalent as to their national economic effect." The *if* may clearly destroy more often than not any determinateness in the process of choice-making. But when it fortunately has not, the saving of labor has to be gauged by no less than five of what are termed "indices." They are: the extent to which use can be made of productive capacities; input of raw materials and fuel and power per unit of output; per-man productivity; finally current costs per unit of output and capital outlay per same unit, i.e., capital intensity of this unit. With ill luck the poor project-maker may find himself saddled with five vectors each pointing in different directions.

However, the authors of the pronouncement are clearly aware that what the project-maker expects from them is an "index" which would uniquely determine the superiority of one alternative over another; hence without saying so explicitly, they seem to hint that of the five indices, current costs and capital outlay per output unit are of higher rank.

Two opposite schools of thought suggested in the Debate that the final test in choice-making should be based on either current costs or on capital outlay per unit of output (the size of output being for the project-maker, of course, a datum derived from the Plan). In

addition, some of the participants in the Debate denied the commensurability of the two criteria: as one writer says — quoting a Russian proverb — you cannot compare a bluebird in hand with a crane in the skies. The approach of either school has been refuted and so has been the thesis of incommensurability. The reasons given for the refutation are more illuminating than the refutation itself.

Those who postulated minimizing capital outlay/output, i.e., those who advocated the least capital intensity of the process of reaching the targets, laid themselves open to the charge that their criterion cut across the path of technological progress. Clearly this could not be accepted as a point of economic policy. On the other hand, those who advocated minimizing current-costs/output ignored the main headache in running the Soviet economy, the chronic and acute shortage of capital. True, they could point out with prima facie good reasons that in the current-costs/output "index" the dividend by definition takes care of capital outlay, since it includes depreciation allowance. However, the pronouncement explains, "depreciation allowance per unit output amounts as a rule to very little," while to the socialist state the absolute magnitudes of capital involved in this or that project are of paramount importance (p. 105). Thus we can hear the skeleton rattling in the cupboard. The point of the over-all demand-supply position of the economy has been touched upon. The dilemma is: how, in establishing preference among different capital-using processes, should scarcity of capital be taken care of in a system in which the scarcity effect on the price of capital is not acknowledged? (Incidentally, along with the inadequacy of conventional amortization charges for the purpose of testing different investment variants the pronouncement points significantly to the fact that "prices of producer goods which go into the current costs express the actual expenditure of social labor very incompletely and to a varying degree" — an admission of relevance far beyond the proper subject of the Debate.)

One of the leading lights has at least sensed the problem. His suggestion was to allow by the use of some rather unwieldy device for the chain effect which a given project would have on capital demand throughout the economy. While otherwise the same author's approach has found little credit in the pronouncement, this particular suggestion has met with qualified approval. Assessment of the chain impact of a project on the total capital demand must have appeared too nightmarish a proposition for the idea to be accepted without qualifications. Yet the authors of the pronounce-

ment proved unable to formulate them. It has failed to tell the project-maker how far he was expected to follow up the impact, or what method and tools were to be applied by him.

Nevertheless — for obvious didactic reasons (see note 4, p. 451) — the authors of the pronouncement while leaving this sore question unresolved, found it necessary to insist on the principle of commensurability of current costs and capital outlay. True it is not easy to see how this commensurability could be questioned on Marxian grounds since both are supposed to be reducible to labor expenditure and consequently to be homogeneous.

As it appears, less concern for orthodoxy than for the practical need of a yardstick has caused the project-makers in the Soviet Union to use widely the unsophisticated device called "coefficient of efficiency" of capital investment:

$$E = \frac{C_1 - C_2}{K_2 - K_1}$$

where the Cs and Ks stand for capital outlays and current costs of two tested alternatives respectively (the clumsy formula would not permit comparisons of more alternatives at a time). In practice different Soviet project-making bodies apply different and differentiated coefficient tables compiled on some purely empirical basis. In various branches of the economy the E varies as widely as from 3 to 25 per cent. Time and again the crude device has been repudiated by authoritative voices, but the harassed project-maker seems to have kept applying it on the quiet. Now it has been formally disowned in the pronouncement because of its embarrassing family likeness to the "economic category" of the capitalist rate of profit.

"Under socialism," the pronouncement proclaims, "what is at issue is not the question of obtaining a given norm of profit but one of the period over which the additional capital outlay should justify itself by the saving on current costs." In other words while the E-device is dogmatically inadmissible, a *nihil obstat* to its use could be granted after some slight treatment with algebraic cosmetics: i.e., if instead of E, its reciprocal is used. Now, the

$$1/E = \frac{K_1 - K_2}{C_2 - C_1}$$

is, of course, nothing else but the familiar recoupment period.

However, whether E or, for reasons of ideological elegance, its reciprocal, is used, it is imperative that:

(a) the capital outlay/current costs substitution allow adequately for the different time pattern of the variants compared. Some of those who took part in the Debate have suggested a discounting procedure on the familiar line by $A(1 + \Delta)^t$ with the Δ having all kinds of meanings — much ingenuity being used by different writers to keep it at as safe a distance as possible from the rate of interest;

(b) the project-maker be told the method of finding a ceiling for the recoupment period.

In either respect the pronouncement refuses to commit itself. It side-steps the issue of methodology for discovering a maximum recoupment period. And it shows itself studiously vague as to the time-impact issue. Only in passing, a hint is dropped as to the advisability of allowing for that impact by the use of "perspective" rather than static price structure. But how is the project-maker to find a clue to "perspective" prices? The discounting formula $A(1 + \Delta)^t$ seems to be implicitly disapproved: even Professor Strumilin's admirable endeavors to supply the Soviet economy with a Δ as respectable as possible, by defining it as a rate of devaluation of capital (i.e., of "embodied" labor in the Marxian concept) through growing "live" labor productivity, has been passed by the pronouncement's authors in significant silence.

Instead the proposition has been expounded that the increment in national income, which results from selecting an investment variant that is less capital-intensive, is positively related to (a) the amount of capital thereby saved; (b) the period over which this amount has been released for other uses; and (c) "the magnitude of national income created in the given year per one rouble of capital assets in operation throughout the country" (p. 112). The most remarkable thing about the proposition is that in so far as it is not self-evident it leads the pronouncement straight into the danger area of a uniform rate of return on capital and a uniform rate of interest (although, as it were, an "ex post" one, with no allowance for expectations).

I do not think that the oracular statement has done any positive harm to the Soviet practice of project-making. In fact the project-making bodies might sigh with relief having obtained a blessing for the $1/E$, at least something to fall back on, clumsy and imperfect as the tool is. As to the theoretically minded economist, it must have left him bemused and bewildered. It has extended the scope of the

dilemma by insisting on the existence of quantitative economic criteria of investment choice, binding the project-maker and the planner alike, and it has almost in the same breath released the latter from such criteria and subordinated his decisions to socio-political imponderables only.   It has tried to offer theoretically coherent guidance, yet it has stopped short of that since it has failed to show the way of finding out how expensive capital is in the Soviet Union, what its equilibrium price is today and is expected to be tomorrow. To the student, however, the pronouncement is of the greatest value, since it lays bare with singular clarity the tremendous difficulties which beset anybody trying to give a coherent solution to the interest problem within a doctrinal framework to which this price is alien.[4]

ALFRED ZAUBERMAN.

LONDON, ENGLAND.

4. One may wonder why the authorities who have assumed an appearance of neutrality during the Debate and in fact seemed to have stopped the Debate for quite a time, have rather unexpectedly come forward with what was intended to be at least a provisional verdict.   The clue may lie perhaps in the admission that hesitations on what appears to be in Soviet project-making the main issue, i.e., the capital-outlay/current costs substitution, have caused in recent years "not inconsiderable harm."   Obviously this harm is the reason why the sleeping dogs have not been let lie.

UNEMPLOYMENT IN PLANNED AND CAPITALIST ECONOMIES: COMMENT*

By F. D. Holzman

In a very interesting article in this *Journal*,[1] Alfred R. Oxenfeldt and Ernest van den Haag compare "real" and "monetary" inflation and depression for centrally planned and capitalist economies respectively. I am in substantial agreement with most of their analysis which constitutes an important contribution to the theoretical literature on comparative economic systems. I have reservations, however, concerning one of their more important conclusions that while real and monetary inflation and depression are interrelated under capitalism, they are independent in planned economies.[2] I should like to indicate:

1. the particular set of assumptions regarding the operation of a planned economy from which this conclusion must have been derived;

2. the reasons for considering these assumptions extreme as well as evidence for thinking the authors had more realistic assumptions in mind when they wrote their article;[3]

3. and finally the change in conclusions which would follow from adopting the more realistic assumptions.

In the discussion which follows, I rely heavily on my observations of the Soviet economy; except where explicitly noted, however, my

1. Alfred R. Oxenfeldt and Ernest van den Haag, "Unemployment in Planned and Capitalist Economies," this *Journal*, LXVIII, 43–60.

2. "While in capitalist economies monetary and real phenomena occur jointly, planned economies may suffer separately from monetary and real inflation and from monetary and real depression. The remedies available to planners for monetary inflation and depression are virtually identical with those available to capitalist economies but they can be used without regard to effects on production" (pp. 59, 60).

Note the following definitions:

Real inflation in a planned economy "occurs when plans call for outputs exceeding what the available resources can actually produce."

Real depression in a planned economy occurs "if output goals were set so low as to leave available resources, goods, or workers unutilized."

3. In all fairness the authors say that they "purposely overdraw the contrast between the price system and centrally planned economies in order to point up differences which reality blurs" (p. 44).

assumptions should not be taken to represent accurately economic conditions in that country.[4]

I

The authors define planning "as a system which replaces the independent and autonomous market mechanisms with government decisions" (p. 43, n. 1). At no point do they spell out their assumptions in greater detail. A planned economy which is consistent both with this quotation and the conclusion quoted above would have to operate as follows. Every transaction at the production level would be accomplished by physical controls. Specifically this rules out the use of free markets for the allocation not only of land, raw materials, and equipment, but also of labor. No reliance would be placed, therefore, on material incentives for the accomplishment of tasks by labor at either the worker or managerial levels; every individual would be told by the central planning board where he was to work, the nature of his task, and presumably, the penalties for failure.[5] Money flows would not be necessary to accomplish transactions at the level of production either in the purchase of land, capital, labor or raw materials; money may, however, serve as a unit of account.[6] Distribution is the only exception to the foregoing; it may be assumed that consumers' choice prevails in the consumers' goods market and is effectuated by the use of currency. Since monetary incentives have been ruled out, however, the amount of money given to each household must be a function of some factor other than the quantity and quality of work performed (e.g., social status, age, need, equality).

If these assumptions are made, the authors' deductions follow logically. Monetary inflation or deflation can occur only in the consumers' goods market since this is the only market in which money functions as a medium of exchange. Money spent in consumers' goods markets has, however, no channel for seeping back to production levels (where money is not used to effect transactions); the level

4. Material cited below which relates to the Soviet economy can be found in the following places: Raymond P. Powell, *Soviet Monetary Policy* (doctoral dissertation), University of California, 1952; F. D. Holzman, *Soviet Taxation* (Cambridge: Harvard University Press, 1955), chaps. 1, 2; *idem*, "Financing Soviet Economic Development" (Part III), *Capital Formation and Economic Development*, in press; Powell's comment on the preceding article and my rejoinder to his comment.

5. We assume here that the central planners *do* care whether or not plans are fulfilled.

6. The authors imply that this is not necessary to their system when they say "no price system guides production" (p. 53).

of output therefore is clearly independent of monetary inflation or deflation in the consumers' goods market. Real inflation or deflation occur at the production level if plans are greater than or less than availabilities. Since money is not used at this level, money cannot affect real output; nor can real inflation or deflation in production have any direct impact on monetary inflation or deflation at the same level. Real inflation or deflation at the production level may cause monetary inflation (deflation) in the consumers' goods market, however, by reducing (increasing) the output of consumers' goods. This is the only connection between the consumers' goods market and the level of output, and between "monetary" and "real" forces.

II

It is difficult to imagine a planned economy operating without labor incentives. Not even the armed forces of modern nations dispense with material incentives. The one planned economy with which we have experience, the Soviet Union, relies as heavily on material incentives as most capitalist nations. In periods of great tension nonmonetary incentives such as patriotism and duty to country can be relied on to a certain degree. In peacetime, the "instinct of workmanship" may serve a similar function on a limited scale. But it is difficult to imagine nonmonetary incentives serving as the driving force of an economy for long under normal conditions. Barring unforseen and fundamental changes in human nature and society, physical control of labor by central planners remains the only comprehensive substitute for monetary incentives. I do not think it is necessary more than to state that, even if workable, physical direction of labor on a national scale would nevertheless be vastly inferior to direction by incentives.[7]

It is also difficult to imagine a system in which money flows are absent at the enterprise level. Unless the purchase and sale of every nail and every brick is voted on by the central planners or their subordinates, a medium of exchange which has some option must be employed. And if money is used for any inter-enterprise transactions, it is highly probable[8] that it will accompany all transactions including those in which it has no option whatever (i.e., those which are com-

---

7. The Soviets attempted substantial equalization of wages in the early period of their development. Although this did not do away with monetary incentives, it sharply reduced them. The effects on productivity, absenteeism, labor turnover, etc., were disastrous. In 1931 Stalin called for a reversal of this policy and highly differentiated wages have been the rule since.

8. Though not absolutely necessary.

pletely subject to direction by the central planners). This is because monetary controls are simpler to administer than physical controls and can be used to support as well as to relieve the burden on physical controls wherever money flows accompany physical transactions.[9]

Although the authors' conclusions imply unrealistic assumptions, it is quite clear that these assumptions were not in their minds when they wrote the body of the article. In several places they refer to "wage-payments" and to "income payments" (pp. 47, 48, 53). The use of these terms, particularly the former, gives the impression that the money paid out to households for the purpose of distributing consumers' goods is an incentive payment for services rendered and not a payment unrelated to work performed.[1] This impression is fortified by the authors' statement (p. 48) that: "higher wages would lead to cost increases and deficits for enterprises." Wage payments which are reflected in enterprise accounts as costs would necessarily be incentive payments to have any significance at all; it would be pointless to keep accounts if wages were determined on the basis of some other criterion.

The authors also imply the existence of managerial incentives. In discussing the attempts of managers to fulfill plan, they say: "They [the managers] could use hoarded resources to overfulfill some of their goals. (The rewards for overfulfillment are relatively high in the U.S.S.R.)" (p. 56).

Finally it is abundantly clear that their planned economy is an economy which utilizes money flows at production, as well as at

9. Money flows accompany all legal transactions at the enterprise level in the Soviet Union. These transactions can be divided roughly into three groups with many gradations between, however. Funded commodities are directly allocated from producers to users. Quota commodities are directly allocated by industry but the amount supplied to an industry can be purchased freely by enterprise within the industry. Decentralized commodities can be purchased and sold freely by all enterprises. The role of money in this spectrum runs from the purely permissive in the case of funded commodities to that of determinant in the case of decentralized commodities, quota commodity transactions falling somewhere in between. (Cf. G. Bienstock, S. Schwarz, and A. Yugow, *Management in Russian Industry and Agriculture* (New York: Oxford, 1944), pp. 58–59).

1. I have one hesitation in submitting this interpretation. In discussing the effects of repressed inflation in a planned economy, the authors do not mention its adverse impact on work incentives. This neglect is, of course, consistent with an interpretation which assumes that monetary incentives are not used in the planned economy. But the authors also neglect to mention this effect of repressed inflation in their discussion of capitalist economies (p. 52 top). Therefore I would guess that the point is one which escaped their attention in treating both economies and that it is not unfair to assume that by wage payments they meant incentive payments.

distribution, levels. The discussion on pages 47 and 48, in particular, leaves no doubts on this score. We have just quoted one relevant passage from page 48. Another passage follows (p. 47):

"Expenditure may be deficient because income recipients save more than the planners expected; or the income distributed, primarily in the form of wages, may have been less than was anticipated; or planned monetary investment outlays may have been set too low . . ."

III

Let us now examine the implications of postulating a planned economy in which money flows accompany all inter-enterprise transactions, and in which there is a free market for labor. It will be shown that monetary and real inflation and depression are, under these assumptions, interrelated at the production level and that inflation and deflation in the consumers' goods and factor markets are functionally related.

Monetary deflation in the factor markets.

Under our assumptions the authorities have to plan monetary flows in the labor, raw materials, and producers' goods markets as well as in the market for final output. Insufficient funds in the factor markets would prevent managers from hiring the workers and purchasing the raw materials and producers' goods necessary to plan fulfillment.[2] This is true regardless of whether the funds held by managers possess infinite option or whether the transactions are directed entirely by physical controls. The result, of course, is that factors of production, including labor, are unemployed and national output declines.

Monetary inflation in the factor markets.

The effect of monetary inflation in the factor markets (due to the availability to plant managers of surplus funds) depends on whether real plans do or do not exceed availabilities and on the structure of managerial incentives. If monetary inflation is superimposed upon real inflation, the results are unambiguously inflationary and will be described below. If, on the other hand, real depression exists, the intensity with which surplus funds are used

2. I am assuming with the authors that wage rates are rigid downward (p. 53) and that prices of raw materials and producers' goods are controlled.

depends largely on the nature of managerial incentives. Suppose, for example, that managerial incentives are structured as follows:

1. rewards for overfulfillment and penalties for underfulfillment of output are high;

2. rewards for overfulfillment and penalties for underfulfillment of profits targets (or money cost targets) are negligible;

3. plans for the year $t + 1$ are not dependent on performance in the year t.

Under this incentive arrangement, one would expect strong competition for labor and a bidding up of wage rates even if real plans are set below availabilities. Managers clearly have everything to gain and nothing to lose from bidding for workers: the rewards for overfulfilling output plans more than offset the penalties for underfulfilling profits if this is a consequence of bidding up wage rates.

Condition (3) is probably not realistic, however. In a developing economy, the central planning board could hardly avoid adjusting the plans of $t + 1$ to the performance of t, particularly if performance in t represents a substantial change relative to $t - 1$ (or if new methods of production are about to be introduced). Plans must, after all, be based on past experience if they are to be meaningful. Under these more realistic conditions and real depression, the availability of surplus funds will not necessarily lead to inflation in the labor market. For fulfillment or overfulfillment of plan in t would lead to higher targets and to the risk of underfulfillment and attached penalties in $t + 1$. The behavior of a plant manager would depend in the final analysis on the relative amounts of the rewards for fulfillment and penalties for underfulfillment of output viewed over the long run.

While we consider conditions (1) and (2) realistic in light of Soviet experience,[3] planners under somewhat different circumstances might well choose to stress incentives to fulfill profits and to minimize output incentives. When profit-emphasizing incentives are combined with real depression one cannot predict on a priori grounds whether factory managers will or will not spend surplus funds. Assuming that output prices are fixed by the state, the profits of enterprise become

---

3. Berliner's interviews with Soviet managers indicate that fulfillment of output goals is much more highly rewarded than fulfillment of profits goals. (Cf. Joseph Berliner, "The Informal Organization of the Soviet Firm," this *Journal*, LXVI, 342–65). For a contrary view see: H. E. Ronimois, "The Cost-Profit-Output Relationship in a Soviet Industrial Firm," *Canadian Journal of Economics and Political Science*, May 1952, pp. 173–83, *Comment* on the above article by the present writer in the same *Journal*, Nov. 1953, pp. 523–31, and *Reply* by Ronimois, May 1954.

a function of two factors: cost and output; the behavior of factory managers will depend on how changes in costs and output affect profits. They will bid for workers if the profits from greater output more than offset the effect of rising costs; if on the other hand the effect of rising costs is to reduce profits faster than they are augmented by increments to output, competition for labor will decline.

In the light of the preceding discussion the effects of monetary inflation in the factor markets would appear to fall into three categories.

(1) If managers have an incentive to use surplus funds because there is real inflation, there would undoubtedly be competitive bidding for workers. This would adversely affect productivity and output for several reasons. First, labor turnover would increase involving a rise in frictional unemployment.[4] Second, in the short run at least, productivity would suffer as workers faced the problem of adjusting to new jobs. Third, there would be hoarding, hence misallocation of labor. Production would be reduced as some firms held more workers than could be productively employed while others were short of needed workers.

(2) If there is real depression but managers have an incentive to use surplus funds for reasons already noted, the results are likely to be salutary. Expenditure by managers of surplus funds will at the very least reduce the level of unemployment and may lead to full or overfull employment.

(3) If there is monetary inflation and real depression, and managers have no incentive to spend surplus funds, the incentive to produce efficiently is blunted. With both monetary and real resources in surplus supply, there would be no positive reason for attempting to reduce costs of production. Real costs would tend to rise and productivity to fall.[5]

4. Wages rose, in the Soviet Union, from an average of 700 rubles per year in 1928 to over 4000 rubles in 1940. This was substantially in excess of the increase called for by successive plans. Competitive bidding by managers for workers which was responsible for the wage increase was also largely responsible for a very high degree of labor turnover. Labor turnover reached a peak in 1930 of more than 150 per cent a year.

5. The incentive to efficiency will be at a peak when monetary deflation and real inflation are paired together. This is because money is scarce in the former and factors of production in the latter. Greater efficiency under these circumstances, by conserving scarce materials, could pave the way to greater output hence greater rewards.

The Soviet literature, incidentally, makes frequent reference to the blunting effect of surplus funds on managerial incentives (cf. Holzman, *Soviet Taxation*, chap. 4).

Monetary inflation in factor and consumers' goods markets.

Monetary inflation in the consumers' goods markets is due, under the authors' assumptions, to either excessive income payments or short-fall in the production of consumers' goods. In an economy in which labor is directly allocated and in which payments to labor are not related to the quantity and quality of work performed, inflation is easy to remedy by, say, raising taxes or perhaps adjusting the basis upon which workers are paid. If, however, labor receives an incentive wage in a free market, monetary inflation in the consumers' goods market is dependent on events in the labor market. An inflation in the market for consumers' goods which originates in the labor market would be more difficult to predict and to control. This is because in planning one would have to reckon not only with the expenditure habits of consumers and the quantity of goods available for sale but with the behavior of both managers and workers in the labor market as well.[6]

If monetary planning in the labor market should provide managers with surplus funds on a continuing basis,[7] a phenomenon resembling capitalist wage-price spiral inflation would develop. In several respects the inflation mechanism would be different, however. The initiative for wage increases, in a planned economy of the type under discussion, would come to a much greater extent from the managers and to a lesser extent from the workers for two reasons. First, if funds are available and if managers are output- rather than profit-oriented they will not feel restrained from bidding up wages competitively even if this should involve monetary losses. Second, since unions presumably would not be allowed to function as wage-determining institutions in a centrally planned economy, the workers could not exert much upward pressure on wage levels.[8] Perhaps the principal significance of this is that central planners need not hesitate to use price-increasing taxes[9] for fear of inducing "effective" wage demands on the part of workers. It should also be noted that an increase in money wages would lead to an increase in real wages only

---

6. The Soviets consistently underestimated the extent to which management would evade controls designed to prevent overspending on wages. The actual average wage exceeded plan throughout the thirties.

7. This has been the case in the Soviet Union for special reasons which cannot be discussed here. Cf. the writings of Powell cited earlier.

8. They would undoubtedly prove not unwilling accomplices in the act, however.

9. This is the type of tax upon which the Soviets rely for most of their revenue. See Holzman, *Soviet Taxation*, chap. 9.

in the unlikely case that the planners followed consumers' sovereignty in the choice of the rate of investment.[1]

Monetary inflation in the consumers' goods market and the level of output.

Given the existence of a free labor market, monetary inflation in the market for consumers' goods, no matter how generated, will have an effect on the level of output through the incentive to work. The effect of repressed inflation is, of course, to reduce work incentives and output. Whether open inflation, and taxes used to offset inflation, increase or reduce the incentive to work depends among other things on the kinds of taxes levied and on the relative importance of substitution and income effects on large and heterogeneous groups of people. Alternative possibilities are too space-consuming and too well-known to be elaborated here.

IV

To summarize: real and monetary phenomena are not unrelated in models of planned economies which are based on the use of money as a medium of exchange and on monetary incentives. In the factor markets real depression may be remedied by monetary inflation if managerial incentives are appropriately structured. Monetary deflation, on the other hand, could cause unemployment even under conditions of real inflation. Moreover, the factor markets and consumers' goods market are interrelated. Inflation, particularly repressed inflation in the latter, will result in reduced output. And finally monetary inflation (deflation) in the consumers' goods market will depend on the combination of real and monetary forces in operation in the factor markets.

Real and monetary inflation and deflation are interrelated, therefore, in both capitalist and planned economies, albeit the interrelationships in the one and in the other do differ. There is no clearcut distinction between the two economies on this ground. A second distinction mentioned by the authors is that in planned economies there are no automatic cumulations of depression and inflation whereas under capitalism cumulations are characteristic phenomena. In planned economies, no "intensification of swings due to changing expectations will occur." This, in my opinion, is much the more basic distinction between capitalist and centrally planned economies.

UNIVERSITY OF WASHINGTON   F. D. HOLZMAN.
SEATTLE

---

1. Since the rate of investment is probably the most important politico-economic decision to be made in a planned economy, consumer sovereignty on this matter can be ruled out as a matter of definition.

REPLY

By ALFRED R. OXENFELDT AND ERNEST VAN DEN HAAG

We are indebted to Mr. Holzman for crystallizing several crucial issues. Owing to space limitations we are obliged to neglect some of the acute points he raises and confine ourselves to two of them.

The first (elaborated in sections I and II of his comment) arises from a misunderstanding of our definition of planning: "a system which replaces the independent and automatic market mechanism with government decisions." Mr. Holzman takes this to mean that "every transaction on the production level would be accomplished by physical controls" and that "no reliance would be placed therefore on material incentives." This is unrealistic, he comments, and "there is evidence" that we had "more realistic assumptions." There must be. Indeed, this *exclusively* physical conception of planning never occurred to us. But obviously we were not clear enough and must apologize to Mr. Holzman and to others who may have been led astray.

We meant to define planning as a system in which the market would not be autonomous, but subordinated to government decisions and used only to help achieve production goals set by the planners. We did not imply exclusively "physical controls" or absence of "material incentives," but only that decisions, whether transmitted physically or through monetary means, are made by the planning authority rather than determined by market forces. The authority may adjust monetary incentives accordingly, for instance, to offer high rewards for plan fulfillment. But these are not effects of autonomous market decisions — on the contrary they are means by which the planning authority carries out its decisions. Hence: "demand, supply and prices depend on the preferences of the planners" (p. 52, n. 1). There is no market *autonomy;* and no more *semi-automatic* interaction of monetary incentives (in practice often less) than is consistent with central determination of outputs by the planning authority.

Having pointed out that to assume planning without money and monetary incentives is rather unrealistic, and then that we perhaps did not assume it, Mr. Holzman in section III of his comment himself assumes that the government transmits its planning decisions *exclusively* through monetary means. Production is directed exclusively through money flows (in our terminology, automatically though not autonomously). Mr. Holzman concludes, as indeed he must, that

monetary disorder would interfere with planned production when money is the exclusive means of insuring compliance with plans. For instance, insufficient money flows would cause what we have called "real depression" because managers would not have the money with which to pay the factors of production needed to carry out plans — even if the plans provide for full employment. We do not dissent from Mr. Holzman's reasoning or conclusion but find his assumption — central planning with exclusively monetary means — unrealistic, indeed paradoxical. It is no less odd than the idea of *exclusively* physical transmission of central plans which we are halfheartedly accused of fathering.

Mr. Holzman's argument reveals how peculiar his premise is. For he assumes that the planning authority continues to rely exclusively on monetary means to achieve the planned outputs, while not providing for sufficient money flows. We assumed instead that should they be insufficient, money flows would be adjusted to plans rather than vice versa. Planners would not rely on money flows where they interfere with production plans: "Prices and aggregate expenditure follow but never determine output decisions" (p. 47). If they do not follow properly, the government would "not elect to reduce output and employment" but rather would increase money flows. To us it seems fundamental to the logic and history of central planning that when money flows hinder output plans, planners will not elect to sacrifice the output plans to the monetary mistakes. Otherwise they would reintroduce the very automatism planning was meant to replace. In practice, why should government banks refuse credit to managers as needed to fulfill output goals? However much automatism they might otherwise retain, the elimination of the automatic relationship in which monetary demand is the main inducement of, and guide for, output decisions it the very *raison d'etre* of planning. To plan and yet to allow deficient money plans to reduce production would indeed be to forego one's cake only to throw it away.

Mr. Holzman's second major point (section III of his comment) in our opinion involves a basic and instructive misconception of central planning. Mr. Holzman asks what would happen if planners "choose to stress incentives to fulfill profit and to minimize output incentives" (p. 464). But under planning, profit incentives must foster, or at least not hinder, achievement of output goals. Possibly there are ways of increasing profit incentives somewhat without defeating output plans. But, contrary to Mr. Holzman, output plan

fulfillment rewards must always be higher than rewards for anything else,[1] unless central planning is to be abandoned.

Usually planned output goals call for production assortments different from those desired by the market. Therefore, there always is an illegal profit incentive to depart from the plan. On the input side, similarly, the manager might reduce costs by procuring the factors he wants on his own, rather than by using the factors and sources planned for him. Soviet literature bears witness that this happens now not infrequently though it is illegal: managers ignore or reshape part of the plan to make illegal profits; or to fulfill some output goals at the expense of others and get the rewards and escape the punishments. Even under existing conditions in the U.S.S.R., the temptation is strong. But, by and large, managers will find it most advantageous to fulfill the main output goals — though they will sometimes resort to illegal means, or even indulge in some nonplanned production on the side.

If planners did what Mr. Holzman suggests, a manager would be able to earn one reward for fulfilling the plan by producing a given assortment, quantity and quality of products. He would also earn a higher reward (not conditioned on output plan fulfillment) if he showed a profit. Where there is a choice — where he cannot reap both rewards (and this is the relevant case) — he will want to reap the higher reward. Mr. Holzman in effect proposes that plans themselves encourage managers to make profit considerations paramount. Output plan fulfillment rewards would be subordinated to rewards for achieving the prescribed, or, the greatest, price-cost difference. Managers would have every economic incentive to violate and ignore output plans. They would be restrained only by legal punishments, which, even together with the preponderance of legally achieved economic rewards have not in the past proved quite sufficient. Clearly, for planners it would be the height of folly to follow Mr. Holzman's advice — which they would rightly recognize as a thinly disguised suggestion to shelve central planning.[2]

1. Or, which amounts to the same, other rewards must be conditioned on plan fulfillment. But in this case we have already shown that they could help little to eliminate "hidden unemployment."

2. Mr. Holzman is quite right, however, in suggesting that his advice would eliminate the basic inefficiency of central planning. That central planning is not a necessary or optimal solution to the problem of unemployment we have tried to show throughout our paper. But in discussing central planning, we cannot attribute to it the problems of a market system as Mr. Holzman does when he makes central planners rely exclusively on money flows; nor can we attribute to central planning the solution of the market system as Mr. Holzman's proposal does here.

There is one conceivable case in which Mr. Holzman's proposal would be consistent with central output planning. If planned output goals are identical with those that would have emerged from an autonomous market, profit incentives even higher than output fulfillment incentives would not tempt managers to frustrate the plan. The highest profits might be found in achieving planned outputs in the planned manner. But it is too much to expect planners to create the correct and harmonious structure of product prices and factor costs. Moreover if they did — if output goals, costs and prices became identical with those that would have emerged from a free market — they would have achieved cumbersomely what would have happened had they not bothered to plan. It is more reasonable, therefore, to assume that planners cannot rely upon market price and cost structures. Their output goals and the incentive structure for their realization must differ from those of a free market. Profit and planned output incentives therefore must pull managers in different directions. If planned output goals are to prevail, profit incentives must be either conditional or lower than output incentives, even though this incentive structure makes inefficient use of resources unavoidable, as we have shown.

Alfred R. Oxenfeldt.

West Hempstead, N. Y.   Ernest van den Haag.

New York University

THE RETARDED ACCEPTANCE OF THE MARGINAL
UTILITY THEORY: COMMENT*

By JOHN P. HENDERSON

Professor Emil Kauder recently presented in this *Journal* a conjecture regarding the relatively late acceptance of the marginal utility theory.[1] In answer to the question, "Why did economics have to wait until 1870 before recognizing the great significance and importance of use value?" he presented reasons based upon interpretations of the history of economic thought which are at odds with most hitherto accepted doctrine. Some comment upon his unusual theories seems to be in order.

Professor Kauder offers two basic reasons for the long delay in the adoption of marginal utility analysis by British political economy. First, he claims that the hegemony of the labor cost theory in British thought in the seventeenth and eighteenth centuries was due to the power and prestige of Calvinist religious theory. "It was no coincidence," he says, "that the members of the Italian-French subjective value school were Catholics and the defenders of the cost theory of value were Protestants."[2] Secondly, he believes that early nineteenth century formulations of utility were hindered by the strong foundations of Ricardian labor theory in academic circles where utility advocates were unable to exert sufficient influence to gain acceptance for their ideas. Jevons, Menger and Walras, on the other hand, "carried weight because they were professors at well-established universities and not men without jobs and out of luck, like Gossen."[3]

I

A number of considerations are neglected in Kauder's interpretation of the part played by religion in the development of economic doctrine.

Implicitly, Kauder defines "acceptance" of marginal utility theory as acceptance in British political economy. Yet he seems to

1. "The Retarded Acceptance of the Marginal Utility Theory," Nov. 1953, pp. 564–75.
2. *Ibid.*, p. 565.
3. *Ibid.*, p. 575.

feel that men like Lattini, Davanzatti, Montanari and Galiani played as important a role in the development of economic theory as did Locke, Petty, and Smith. He argues that the latter, being Protestants and the intellectual successors of Calvin, set the course of economics in the wrong direction by placing labor "in the center of their theology." He finds greater significance in the fact that some of the early advocates of utility theory were Catholics than in the fact that they were predominantly French and Italian thinkers who dealt with totally different queries than those which concerned British political economy. Thus to emphasize French and Italian thinkers, and to associate the development of economic theory with those economies, would appear to be in contradiction with the general recognition of the fact that political economy as a *discipline* grew hand in hand with British economic development. Britain was the first country to face the problems of transition from a "natural" economy to one in which exchange and the division of labor were dominant. In the process of industrial development, a theory of production and the search for an objective theory of value were necessary preliminaries to the solution of problems which called for policy recommendations.[4] In view of these facts, it would seem that the adherence to labor cost doctrine had a great deal more to do with the problems of economic development than with the religious convictions of its exponents.

Kauder assumes that, had Adam Smith read his Aristotle more religiously than he did his Calvin, he would have realized that the great philosopher's "traditional theory of value" was the correct corner stone of economic science. Thus, he believes, political economy might have been saved from a fateful error that was not corrected until 1870. But the question arises as to just what was Aristotle's "traditional theory of value." Marx claimed that it was an exchange theory and therefore closer to the labor cost doctrine than to notions of utility and individual use value.[5] It must be remembered that Aristotle presented two sides to the value problem, value in use and value in exchange. If Kauder says the former is the most important, then he must at least recognize the existence of the latter and show the way in which the two can be reconciled.

Even granting the conjecture that labor cost theory is Calvinist

4. Speaking of the bullionist controversy, Ricardo remarked, "It could not be settled satisfactorily, because there was no standard to which reference could be made to ascertain whether gold had risen or paper fallen in value." *Works and Correspondence of David Ricardo*, ed. P. Sraffa, VII, 7.

5. *Capital* (Kerr Edition), I, 68–69.

in origin, how does such a formulation account for the two most important advocates of the labor theory of value, Ricardo and Marx, neither of whom was a Protestant? Should not a religious interpretation of the genesis of labor value theory give far more consideration to the intellectual roots of Judaism, than to Calvinism? Nor does the fact that both Ricardo and Marx rejected their parental faith allow Kauder to rule out the influence of Judaism since he claims that "the point is that early education leaves its permanent impression on our minds, regardless of how we may change our convictions at a later date."[6]

However, far more puzzling to a student of the history of economic thought is Kauder's linking of Aristotelian-Thomistic doctrine with utility theory rather than with early formulations of the labor cost theory. Scholars of the history of economic ideas have consistently associated canonist doctrine with the foundations of labor theory and the notion of just price, in which both the buyer and the seller concern themselves with society's goals and needs. Speaking of value, Bernard W. Dempsey, one of this country's leading Roman Catholic clerical economists, says that it "rests upon a kind of estimate, not of the buyer and seller alone, but of the whole community . . . and exchange must take place according to the community's estimate of the social utility . . . because the producer who expects sustenance from society in return for his labor, by performing his function in the social organism, has earned his right to a just price . . . for when society permits transactions at other than just prices, it is cutting off its nose to spite its face. . . ."[7] This hardly seems consistent with the individualism of Carl Menger, who claimed, " . . . where the available quantity does not suffice for all [the general rule of economics], every individual will attempt to secure his own requirements as completely as possible to the exclusion of others."[8]

At issue in this contrast between the Catholic doctrine of just price and the amoral theory of Menger and the utility school is the distinction between what is and what ought to be. The whole base of economics shifted with the development of utility analysis, which postulates that "economics deals with ascertainable facts; ethics, with valuations and obligations. The two fields of inquiry are not on the same plane of discourse. . . . Propositions involving the verb 'ought'

6. Kauder, *op. cit.*, p. 566.
7. "Just Price in a Functional Economy," *American Economic Review,* (Sept. 1953), pp. 482, 486.
8. *Principles of Economics*, trans. and ed. by James Dingwall and Bert F. Hoselitz (Glencoe, Illinois: The Free Press, 1950), p. 97.

are different in kind from propositions involving the verb 'is'. And it is difficult to see what possible good can be served by not keeping them separate, or failing to recognize their essential difference."[9]

How utility analysis could have developed from the Catholic doctrine is difficult to see in view of this basic philosophic difference in purpose and method. Furthermore, Father Dempsey's viewpoint seems in accord with this interpretation for he says, "Apart from the Marxists, the last group of economists who sought to deal with economic and moral questions in theory as they are in fact, identified in an objective institution or situation, were the scholastic moral economists, whose work reached its greatest development in the seventeenth century."[1] But Kauder claims that, during the seventeenth and eighteenth centuries, Catholic writers were correctly advancing the principles of utility and had solved all of its problems except one. What that one was, Kauder does not make clear, but certainly Catholic writers have never yet indicated sufficiently a way in which notions of utility can be reconciled with a theory of just price.

Dempsey is not alone in interpreting Catholic economic thought as a macroscopic, socially oriented doctrine. Another leading clerical economist, Richard E. Mulcahy, in discussing the contribution of Heinrich Pesch, who formulated a complete system of economic theory based upon Aristotelian-Thomistic philosophy, says, "Like Adam Smith, Pesch conceived of economic activity *as aiming at something objective, in a physical sense. It is not a question of pain or pleasure, or satisfaction in a psychological sense, but a provision for the material needs of the people*" (italics added).[2]

The outright contradictions between Kauder's conjecture and the fundamentals of Catholic economic thought appear to stem from his failure to emphasize the development, beginning with Aristotle, of two theories of value, dealing separately with value in use, conceived as subjective and personal, and value in exchange, which was conceived as being the objective result of labor. Calvinist doctrine did not give birth to the notion that labor was important to production. The eminent economic historian, W. J. Ashley, has said of canonist doctrine, "Labour . . . as the one element in production

9. Lionel Robbins, *The Nature and Significance of Economic Science*. pp. 148–49.

1. "The Ability to Pay," this *Journal* (May 1946), p. 353.

2. "The Welfare Economics of Heinrich Pesch," this *Journal* (Aug. 1949), p. 343.

which depended upon human will, became the centre of their doctrine. All wealth was due to the employment of labour on the materials furnished by nature; and only by proving that labour had been engaged in bringing about the result could the acquisition of wealth by individuals be justified . . . the doctrine had thus a close resemblance to that of modern socialism: labour it regarded as both the sole (human) cause of wealth, and also as the only just claim to the possession of wealth. It differed from modern socialist teaching only in allowing different kinds of services to society to be remunerated at different rates. Yet it would certainly have visited with a similar condemnation the acquisition of wealth by sheer speculation, or by the manipulation of the market. The influence of this conception of labour is seen in every part of the canonist doctrine."[3] Yet Kauder states that Catholic doctrine had " . . . no compulsion to integrate labor costs into the social order, or into the philosophy of economic value."[4] The question is, what did integrate the social order if it was not labor? What was just price to be based upon, in the objective sense in which Mulcahy speaks? If we accept the usual interpretation of the canonists, then the true descendent of Aquinas is not Menger, or Jevons; on the contrary, as another eminent economic historian has suggested, "the last of the schoolmen was Karl Marx."[5]

One further point must be made about Kauder's religious interpretations. Up to the present time the usual view of Calvinist doctrine has been that it was associated with the rise of individualism and the justification of money-making, profit, and interest. Such a view not only makes no allowance for the labor theory of value, but seems highly inconsistent with it. Kauder offers no reinterpretation of Calvinist doctrine which substantiates his claim that Calvinist Protestantism nourished seventeenth and eighteenth century cost theory. He makes no attempt to deal with such statements as Tawney's: "It is not wholly fanciful to say that, on a narrower stage but with not less formidable weapons, Calvin did for the *Bourgeoisie* of the sixteenth century what Marx did for the proletariat of the nineteenth, or that the doctrine of predestination satisfied the same hunger for an assurance that the forces of the universe are on the side of the elect as was to be assuaged in a different age by the theory of historical materialism."[6]

3. *An Introduction to English Economic History and Theory* (New York, 1912), pp. 393–94.
4. Kauder, *op. cit.*, p. 569.
5. R. H. Tawney, *Religion and the Rise of Capitalism* (Mentor Edition), pp. 38–39.
6. *Ibid.*, p. 99.

II

In explanation of the utility revolution of 1870 Kauder points out that religion no longer provided the key to economic theorizing. He says that, with the exception of Alfred Marshall, who was still plagued by religious convictions, late nineteenth century economists rejected any serious consideration of costs and supply. Marshall, however, was "attracted to the cost problem," because of his religious bias. This bias was responsible for "the unbalanced character of his [Marshall's] price and value theory."[7] But such a bias, Kauder says, was no longer usual among economists. He offers no explanation of why economists in general had ceased to think in "accordance with religious backgrounds," but claims that the reasons for the acceptance of marginal analysis "can be found only in the history of economic science itself." A scientific attitude, free of religious fetters, enabled British economists to accept marginal utility theory, and the academic prestige of Jevons, Menger, and Walras was the necessary and sufficient means of implementing that acceptance. The implication is that economic theory is independent of events in the economic world and that the development of marginal analysis was unrelated to the changing conditions of English society.

Many questions arise in an investigation of these claims. It is not at all clear in what sense Kauder considers Marshall's theory to be "unbalanced." Marshall's major contribution to the history of economic thought is usually interpreted as a synthesis of classical cost theory and the newer utility theory. Balance between supply (cost) and demand (utility) is the essential feature of his *Principles of Economics*. It would, of course, be a *tour de force* to demonstrate that the great balancer was guilty of one-sidedness, but it would appear that it is not Marshall's but Kauder's analysis which is unbalanced. Kauder does not seem to distinguish between the development of utility theory in the 1870's and the general marginal productivity theory of the 1890's.[8] The transition from Ricardian classical theory to the Marshallian synthesis of marginal analysis, was marked off by two developments in thought. In the 1870's, Jevons, Walras and Menger shifted the emphasis to demand and explained price determination in hedonist psychological terms, ignoring the cost of production analysis of the classicists.[9] The criticisms

7. Kauder, *op. cit.*, p. 570
8. G. Stigler, *Production and Distribution Theories*, chap. I.
9. For a discussion of the influence of contemporary criticism upon the utility school, see J. Viner, "The Utility Concept in Value and Its Critics," *Journal of Political Economy* (2 parts) 1925, pp. 369–87, 638–59.

of this restricted explanation of price, however, gave way to the recognition that a supply analysis was essential, though not pre-eminent, and marked the beginning of the second aspect of the evolution of the "new economics." This was to culminate in the marginal productivity doctrines of the 1890's, which represented the acceptance of disutility along with utility as one of the two determinants of price. The attempt to ignore supply, as the triumvirate had done, was altered by the development of a subjective and psychological approach to cost. It is to the second stage of this development that Marshall belongs, and his contribution is the recognition that it is not possible to explain price without taking into consideration the explanation of supply. The determinant of the latter, from the standpoint of the individual firm, was essentially dependent upon an analysis of cost. Kauder, if I understand him correctly, would have us believe that utility is all that is necessary to explain price. Marshall would be the first to agree to this, provided one is concerned with only the short period market, but for Marshall, and for most economists, market period analysis is but one of the problems with which price analysis deals.

As for the significance Kauder attaches to the academic prestige of Jevons, Menger, and Walras, one questions the relevance of his argument. He states that the marginal triumvirate gave utility analysis a foothold in the universities, causing the subsequent rejection of Ricardo's "wrong-headed" theories. Just how Ricardian theory got into the universities, where it had to be overcome, is not clear, for its founder worked out most of his theory in the stock market and in Parliament and, with the exception of occasional social visits to Malthus, was in no way connected with any university.

Kauder's thesis seems to rest primarily upon the strength of his claims that alternative explanations for the acceptance of marginal utility theory are in error. In favor of his own interpretation he dismisses (1) the rise of sensualism as suggested by Northrop[1] and Myrdal,[2] (2) W. Stark's discussion of the Kantian influence upon Austrian economics,[3] and (3) the Marxian explanation for the rise of marginal analysis.

The first of these, he admits has some credibility. The second, he rules out for reasons which seem hardly convincing. They are

1. F. S. C. Northrop, *The Meeting of East and West* (New York, 1946), pp. 131 ff.
2. G. Myrdal, *The Political Element in the Development of Economic Theory*, trans. Paul Streeten, p. 125.
3. W. Stark, *History of Economics* (New York, 1944), pp. 3 ff.

based upon a statement attributed to an obscure author, Heinrich Ritter von Srbik, to the effect that Kant was not an important influence in Austrian thought in the nineteenth century because of the dominance of Catholic doctrine in Vienna after 1800. Does this not contradict Kauder's own contention that religious postulates had ceased to influence economic science in the nineteenth century?

To the third, the Marxian explanation, Kauder grants no credence at all. He claims that economists did not become predominantly antilabor and probourgeoisie, and that marginal utility gained precedence independently of political and economic history. Unfortunately for Kauder's argument, the marginalists themselves appear to claim otherwise.

Professor J. M. Clark says the "key fact" in J. B. Clark's theories is "that his statements are oriented at Marx, and are best construed as an earnest, and not meticulously qualified, rebuttal of Marxian exploitation theory . . . he concluded that distribution on this basis (Clark's) is not robbery, as Marxian socialism contended, but is basically honest. This causal concept and especially the ethical conclusions, have been subject to more criticism than any other features of Clark's system; yet he would appear to have regarded them as his most basically important contribution."[4]

As to the motivation of Jevons, it is clear that social issues were very much in the forefront of his thinking. Ten years before his scientific treatise he claimed that the greatest danger to society was "that our working classes, with their growing numbers and powers of combination, may be led by ignorance to arrest the true growth of our liberty, political and commercial . . . for even if all could combine with equal ease they would only make things dear and hinder the production of the commodities upon which we all live."[5] This problem of the growth of the power of unions was to be solved by formulating a new theory of value in which "labor once spent has no influence on the future value of any article: It is gone and lost forever. In commerce bygones are forever bygones; and we are always starting clear at each moment, judging the values of things with a view to future utility . . . labor is never the cause of value. . . ."[6]

The intellectual foundations of marginalist thought and the

4. "J. M. Clark on J. B. Clark," in *The Development of Economic Thought*, ed. H. M. Spiegel (New York: John Wiley and Son, 1952), p. 610.

5. Comments from Jevon's introductory lecture at Manchester. Quoted from T. W. Hutchison, *A Review of Economic Doctrines, 1870–1929* (Oxford: Clarendon Press, 1953), p. 46.

6. W. S. Jevons, *Theory of Political Economy* (4th ed.), pp. 164–65.

reasons for its acceptance are well expressed by Friedrich von Wieser, who indicated thus the reasons for his acceptance of the theory of Menger:

"Never has there been as deep and pressing a need of economic theory as in the present. At the time when the representatives of the classical school were thinking out their principles of economy, practical interest was confined to the question of the degree of freedom which the state should give to private economy as such. The socialistic thinkers fought the laissez-faire attitude of private management. One who gave no weight to their objections set himself scientifically apart from them. One did not dream that the day would come, was indeed near, when the proletariat should be strong enough to give pertinence to its demands by force. . . . Almost everywhere in Europe the proletariat has come forward with such strength that it must be considered and a counter-reform of the economic order proposed. . . .

"*The final distinction between possessors and proletariat will not be successful without the aid of theory.* Both classes have sought scientifically to strengthen that position which their own interests have led them to take, and both have thus made errors with serious consequences. The proletariat thinkers have fastened on untenably impractical dogmas. The bourgeois thinkers on the one hand have developed their own case, but on the other hand have left too many loopholes in the explanation. . . . Power will also make its influences felt. In time, however, theory will be called upon. This time will come so soon as it is realized that raw power can work evil but cannot heal it. But when this happens, theory must have unified its thought sufficiently so that men may find the necessary help in it when once their thoughts advance to a point at which they may subject their wishes to the insistent commands of reality.

" . . . the contradictions which the classical-socialistic theory could not solve, have been removed by the theory of marginal utility . . . it enables us to understand, the sense of the economy may be perverted from a social point of view when the mighty utilize their superiority for themselves."[7]

Kauder seems to imply that both science and religion are intellectual processes taking place in a vacuum. His central methodological position, that only the history of "economic science itself" can shed light upon the acceptance of utility theory, and his emphasis upon religion as the controlling influence upon the history of economic science, are theories which do not stand up under the weight of the evidence that social, political, and economic forces shaped the history of economic science and were directly related to the development of both the labor theory of value and utility analysis. His Procrustean argument is indeed "only conjecture."

JOHN P. HENDERSON.

UNIVERSITY OF MARYLAND

7. *Social Economics*, trans. A. Ford Hinrichs (New York 1927), pp. xviii–xx.

REPLY

By EMIL KAUDER

Mr. Henderson raises essentially three objections against the ideas presented in my paper "The Retarded Acceptance of the Marginal Utility Theory."[1]   He charges me with: misinterpretation of canonist thinking, abortive attempts at refuting alternative explanations, and the use of a strange method.   Let us take up these three points in order.

He asserts that "scholars of the history of economic ideas have consistently associated canonist doctrine with the foundations of labor theory. . . ."   Such scholars are very rare.   Neither Joseph Schumpeter nor Father Bernard W. Dempsey belong to this group. Schumpeter in his monumental *History of Economic Analysis* and Father Dempsey in his fine book *Interest and Usury* claim that neither Aristotle, nor medieval, nor late scholasticism formulated any kind of a labor value theory.[2]

"Kauder offers no reinterpretation of Calvinist doctrine which substantiates his claim that Calvinist Protestantism nourished seventeenth and eighteenth century cost theory," says Henderson.   I do not need it, because Max Weber in his paper on "Capitalism and Protestant Ethics" has marshalled all the essential facts for his claim that Calvinist theology considered incessant labor necessary for attaining the certainty of salvation.[3]   I consider it very difficult to dislodge Max Weber's well-documented interpretation.   Besides I cannot find that Henderson's quotation from Richard H. Tawney speaks against Calvinist glorification of labor.   Tawney, although he

---

1. This *Journal*, Nov. 1953.

2. Joseph A. Schumpeter, *History of Economic Analysis* (New York, 1954), pp. 62, 93. According to Schumpeter and others, the only schoolman who presented a cost theory but not a labor value theory of the just price was Duns Scotus (*Labores et expensae*).   "The late scholastics, particularly Molina, made it quite clear that cost, though a factor in the determination of exchange value (or price), was not its logical source or 'cause'." *Ibid*, p. 98.   "Still less than a cost theory of value can a labor theory of value be imputed to them [i.e., the late scholastics, particularly Molina] though this has been done." *Ibid.*, p. 98, n. 23. Bernard W. Dempsey, *Interest and Usury* (Washington, 1948), pp. 150–54. See also Emil Kauder, "Genesis of the Marginal Utility Theory," *Economic Journal*, LXIII (Sept. 1953), pp. 639–41.   Here ample documentation and literature can be found. See also Raymond de Roover, "Scholastic Economics: Survival and Lasting Influence from the Sixteenth Century to Adam Smith," this *Journal*, May 1955, pp. 168–69.

3. Max Weber, *The Protestant Ethic and the Spirit of Capitalism*, trans. Talcott Parsons, pp. 105 *seq.*

is rather critical of Max Weber, does not attack this part of Weber's theory.[4] Tawney even quotes a letter of Calvin in which the great reformer writes: "Whence do the merchant's profit come, except from his own diligence and industry."[5]

So my interpretation only sums up the results of sound scientific research produced during the last fifty years. I remain still "on the side of the angels," if I emphasize the importance of the earlier Italian economists, who as sound Aristotelians developed the value-in-use theory. Schumpeter, for one, thought very highly of them.[6]

I cannot agree with Henderson, when he writes that it is generally recognized that "political economy as a *discipline* grew hand in hand with British economic development" because Great Britain, so he continues, was the first country which emerged from the natural economy. An Italian economist could easily turn Henderson's argument around and write: "Italy was the first country to face the problems of transition from a 'natural' economy to one in which exchange and the division of labor were dominant." Already in the early Middle Ages, Florence and Venice were great trading posts and production centers with established gold currencies; in Venice for the first time double-entry bookkeeping was described, etc. Therefore, our fictitious Italian author could conclude that Italian economists must have a higher rank than English and American writers. The quality of economists cannot be measured, however, by the nature of the economic life which surrounds them, but only by the originality and truthfulness of their analysis. If one accepts this yardstick of scientific greatness, then Davanzati and Galiani must rank rather high.

My analysis of the reasons for the retarded acceptance of marginal utility theory in the nineteenth century was preceded by a criticism of other explanations, to which my critic takes exception. He does not think that I refuted Stark's theory of the Kantian influence upon Austrian economics. I wonder whether he is aware of the fact that Stark did not give a shred of documentary evidence for his thesis. That I based my statement on the authority of the "obscure" Heinrich Ritter von Srbik is due to the fact that Srbik was during his lifetime the recognized dean of Austrian historians and had an unsur-

---

4. R. H. Tawney, *Religion and the Rise of Capitalism*, pp. 316–17. Here Tawney criticizes Max Weber very thoroughly without touching the glorification of labor.

5. *Ibid.*, pp. 105, 109, 114.

6. *Op. cit.*, pp. 292–93, 300–2 and *passim*.

passed knowledge of Austrian life, culture, politics, and ideas.[7] By the way, I did not write that religious postulates were influencing Austrian economic thinking. Rather I said that there is some reason to believe that Aristotelianism, neopositivism, and Leibnitz had somehow helped to form the Austrian method of economic thinking.

Henderson offers a number of well-chosen quotations to show that Wieser, Jevons, and John Bates Clark were eager to fight Marxism with the weapons of marginalism. He appears to consider these quotations a proof for the correctness of the Marxian economic interpretation of history, but I do not. These authors claimed that marginal utility is better suited to avoid contradictions than "classical-socialist" philosophy, and is a weapon in the fight against socialism. But this has nothing to do with the Marxian exposition of ideological bias. Henderson's quotations would fit into the Marxian pattern, if any one of the writers had glorified the interests and actions of the ruling class and thus drawn up a theory strongly at variance with truth. But nothing like this can be found.

Nicolai Bukharin, the outstanding Marxian critic of the Austrian school, must have been aware of the fact that the glove did not fit the hand; therefore he changed his line of attack. Marginal utility, as a theory of consumption, so he wrote, reflects the goals of decadent capitalists who are fond of cigars, cars, and other things which make life agreeable. Whether this picture of the ruling class is more than a caricature, is not my business to investigate here. That the words quoted by Henderson cannot be fitted into Bukharin's framework, is all that matters. These marginalists may have been eager, perhaps too eager to use their newly forged tools for the criticism of Marxism, yet we have no logical warrant to assume that they were advertising agencies for the consumption habits of the decadent bourgeois class.

It seems to me that marginal utility is ideologically neutral. Furthermore, marginalism has no close connection with any special historical economic situation. The utility concept has been based on general psychological observations which theoretically could be made at any time by any economist. Furthermore, the relationship between labor value theory and the economic life of a special historical period can easily be overrated. The classicists were seeking a yardstick to

---

7. Heinrich Ritter von Srbik, 1878–1951. Full professor at the University of Vienna 1922–1945, member of many learned societies, e.g., Royal Historical Society, London; Société d'Histoire moderne, Paris. Selections from his main works: *Metternich, der Staatsmann und der Mensch* (Munich, 1925); *Das Österreichische Kaisertum und das Ende des Heiligen Römischen Reiches* (Berlin, 1927).

measure exchange rates. The search for such a measuring rule is independent of any particular economic situation. I cannot agree with Henderson, who claims that the search for an objective theory of value was a necessary preliminary "to the solution of problems which called for policy recommendations." I do not know any economic problem of classicism which needed for its solution the strict adherence to the labor theory of value. On the contrary, it seems to me that wherever Smith, Ricardo, or Marx used the labor theory, confusion and inconsistency were introduced.

I certainly agree that "political, social, and economic forces" (Henderson) may have some influence on the development of economic thinking. Yet I do not assume, as Henderson apparently does, that a close causal connection exists between these forces and the history of economic thinking. I wonder why, in the enumeration of these forces, no place is left for religion, philosophy, and ethics. It is my thesis, which according to Henderson is "at odds with most hitherto accepted doctrine "that economic thinking like any other form of cultural and scientific activity, is correlated with philosophical, religious, and ethical convictions.

That my method according to Henderson is unusual, is no objection. New insights have sometimes been gained by deserting worn-out trails. Yet I cannot even boast of having used a new method; I am a late-comer. Max Weber, Talcott Parsons, Schumpeter in his fine analysis of Karl Marx, and Gunnar Myrdal, all have worked on similar lines. I agree with Henderson that the application of my method results only in conjectures. But every interpretation in the realm of history is only a conjecture which can be superseded by a better one, and which must be abandoned in the face of valid criticism.

I regret that I find myself so little able to accept Mr. Henderson's criticisms, but am grateful that he gave me the opportunity further to elucidate some points of my thesis.

EMIL KAUDER.

ILLINOIS WESLEYAN UNIVERSITY

THE SCHUMPETER PRIZE FUND

A Fund of approximately $12,000 has been contributed by friends, colleagues and former students of the late Professor Joseph A. Schumpeter, to be known as the Schumpeter Prize Fund and to be administered by the President and Fellows of Harvard College. Prizes of not less than $1,000 will be awarded periodically from the income of this Fund for original work in the field of economics, each such award to be known as a Schumpeter Prize.

The awards will be made by a Board consisting of three members of the Department of Economics at Harvard University, and will be made in principle once every two years. If the Awarding Board finds that sufficiently meritorious work has not come to its attention during any given two-year interval, it may postpone the awarding of a Prize for a maximum of two additional years. Conversely, when sufficient income has accumulated and work of sufficient merit has been found, the Board may award Prizes at one-year intervals.

Anyone who has made an original contribution to economics is qualified to receive the Prize, regardless of academic standing, affiliation or lack of affiliation with educational institutions, nationality, or any other possible restrictions; provided, however, that this contribution falls within those branches of economics in which the late Professor Joseph A. Schumpeter made his most creative contributions. For the present, the study of economic development, and its interrelations with historical, social, and political evolution, and the interrelations between economic theory, statistics, and history seem to be most appropriate, but relevant theoretical work is to be included.

The primary test in awarding the Prize is to be the depth and originality of perception. Whether the work is an article in a periodical or a full-length monograph is irrelevant. Preference will be given to original work published since the awarding of the last previous Prize, but in the discretion of the Board the Prize may be awarded for work published at an earlier date.

There are no age limitations on the potential recipient of the award, though preference will be given to economists between the ages of 25 and 35 years, in line with Professor Schumpeter's reference to "thought and struggle during the third decade of life, which in the case of every thinker, creates what is subsequently worked out."

If on the basis of experience it appears that the foregoing provisions are too restrictive, the Awarding Board may revise the qualifications.

The Awarding Board makes the following announcement:

"The first award of the newly created Schumpeter Memorial Prize has been made to Professor Ragnar Frisch of the University of Oslo, Norway, in recognition of his numerous path-breaking achievements in the field of economic theory, statistics, and econometrics. Schumpeter's own admiration for his genius and leadership is shared by all those who appreciate the use of rigorous, scientific methods in the social sciences."

June 1955

RECENT PUBLICATIONS

Adams, Leonard P. and Mackesey, Thomas W. Commuting Patterns of Industrial Workers. Ithaca, New York: Cornell University, Housing Research Center. 1955. pp. 135. $2.00. [Research Publication No. 1.]

The American Assembly. United States Agriculture: Perspectives and Prospects. New York: Graduate School of Business, Columbia University. 1955. pp. 122.

Arab Information Center. The Arab League: Its Origin — Purposes — Structure and Activities. New York: The Arab Information Center. 1955. pp. 31. [Information Papers No. 1.]

——. Basic Documents of the League of Arab States. New York: The Arab Information Center. 1955. pp. 40. [Document Collections No. 1.]

Auboin, Roger. The Bank for International Settlements, 1930–1955. Princeton, New Jersey. International Finance Section, Department of Economics and Sociology, Princeton University. 1955. pp. 38.

Belshaw, Cyril S. In Search of Wealth. Vancouver, British Columbia, Canada: University of British Columbia. 1955. pp. 84. $2.00.

Bendix, Reinhard. Bureaucratization in Industry. Berkeley, California: Institute of Industrial Relations, University of California. 1955. pp. 11. [Reprint No. 68.]

Beyer, Glenn H. and Partner, James W. Marketing Handbook for the Prefabricated Housing Industry. Ithaca, New York: Housing Research Center, Cornell University. 1955. pp. 64. [Research Publication No. 2.]

Black, John D. Farm and Other Operating-Unit Land-Use Planning. Cambridge, Massachusetts: Harvard University, Seminar in Land-Use and Conservation. 1955. pp. 119.

Blodgett, Ralph H. Our Expanding Economy: An Introduction. New York: Rinehart and Company, Inc. 1955. pp. 973. $6.00.

Brown, Lyndon O. Marketing and Distribution Research. New York: The Ronald Press Company. 1955. pp. 561. $6.75. [Third Edition.]

Chamberlain, Neil W. A General Theory of Economic Process. New York: Harper and Brothers. 1955. pp. 370. $5.00.

Chernick, Jack. A Guide to the Guaranteed Wage. New Brunswick, New Jersey: Institute of Management and Labor Relations, Rutgers University. 1955. pp. 50. [Bulletin No. 4.]

Congard, Roger-Paul. Etude Econometrique de la Demande de Tabac. Paris: Librairie Armand Colin. 1955. pp. 99.

Dowd, Herbert Walter. Non-White Land and Labor Policies in South West Africa from 1918 to 1948. Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 7. [Thesis Abstract, 1954, No. 2.]

Edwards, Charles Pastene. The Western League Powers and the Italo-Ethiopian Affair. Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 7. [Thesis Abstract, 1954, No. 8.]

Ernst, David Howie. Civil Aviation in the Arab League States. Medford, Massachusetts. The Fletcher School of Law and Diplomacy. 1954. pp. 7. [Thesis Abstract, 1954, No. 3.]

Fabricant, Solomon. Government in Economic Life. New York: National Bureau of Economic Research. 1955. pp. 78. [Thirty-fifth Annual Report.]

Feer, Mark Cecil Iselin. India's China Policy, 1949–1954. Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 9. [Thesis Abstract, 1954, No. 4.]

FRANCE, ROBERT R. Union Decisions in Collective Bargaining. Princeton, New Jersey: Industrial Relations Section, Department of Economics and Sociology, Princeton University. 1955. pp. 49. $2.00.

FRIEND, IRWIN; HAMBURG, MORRIS; SCHOR, STANLEY. Pricing and Price Differentials on Over-the-Counter Markets. Philadelphia, Pennsylvania: University of Pennsylvania Press. 1955. pp. 119. $1.25.

GALBRAITH, JOHN KENNETH. The Great Crash, 1929. Boston, Massachusetts: Houghton Mifflin Company. 1955. pp. 212. $3.00.

GINZBERG, ELI (EDITOR). What Makes an Executive? New York: Columbia University Press. 1955. pp. 179. $3.50.

GOORMAGHTIGH, JOHN. European Coal and Steel Community. New York: Columbia University Press. 1955. pp. 65.

HAIRE, MASON. Role-Perceptions in Labor-Management Relations: An Experimental Approach. Berkeley, California: Institute of Industrial Relations, University of California. 1955. pp. 216. [Reprint No. 66.]

HARWOOD, E. C. Reconstruction of Economics. Great Barrington, Massachusetts: American Institute for Economic Research. 1955. pp. 112. $1.00.

HERMAN, FREDERICK. The Victors and the Vanquished. Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 7. [Thesis Abstract, 1954, No. 5.]

HICKMAN, BERT C. The Korean War and United States Economic Activity, 1950–1952. New York: National Bureau of Economic Research. 1955. pp. 64. 75 cents. [Occasional Paper 49.]

HIGGINS, BENJAMIN AND MALENBAUM, WILFRED. Financing Economic Development. New York: Columbia University Press. 1955. pp. 339. 25 cents.

HOLLY, J. FRED AND MABRY, BEVARS D. Protective Labor Legislation and Its Administration in Tennessee. Knoxville, Tennessee: University of Tennessee Press. 1955. pp. 216.

HSIA, RONALD. Economic Planning in Communist China. New York: Institute of Pacific Relations. 1955. pp. 89.

JACKSON, ELMO L. The Pricing of Cigarette Tobaccos: A Study of the Process of Price Development in the Flue-cured and Burley Auction Markets. Gainesville, Florida: University of Florida Press. 1955. pp. 239. $4.75.

KENNEDY, VAN DUSEN. Grievance Negotiation. Berkeley, California: Institute of Industrial Relations, University of California. 1955. pp. 11. [Reprint No. 70.]

KRAUSE, WALTER. The International Economy. Boston, Massachusetts: Houghton Mifflin Company. 1955. pp. 434. $5.50.

LaBASSE, J. Les Capitaux et la Région. Paris: Librairie Armand Colin. 1955. pp. 532.

LEE, LUKE TSUNG-CHOU. The Development of the Consular Institution, 1932–1952. Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 7. [Thesis Abstract, 1954, No. 1.]

LEE, MAURICE W. Economic Fluctuations: An Analysis of Business Cycles and Other Economic Fluctuations. Homewood, Illinois: Richard D. Irwin, Inc. 1955. pp. 633. $6.00.

LIPSET, SEYMOUR MARTIN AND BENDIX, REINHARD. Ideological Equalitarianism and Social Mobility in the United States. Berkeley, California: Institute of Industrial Relations, University of California. 1955. pp. 20. [Reprint No. 76.]

LOMBARDINI, SIRO. Fondamenti e Problemi dell'Economia del Benessere. Milan, Italy: Dott. Antonino Giuffrè. 1954. pp. 128.

MALM, F. THEODORE. Hiring Procedures and Selection Standards in the San Francisco Bay Area. Berkeley, California: Institute of Industrial Relations, University of California. 1955. pp. 21. [Reprint No. 64.]

MATTILA, JOHN M. AND GAUMNITZ, ERWIN A. An Econometric Analysis of Construction. Madison, Wisconsin: Bureau of Business Research, University of Wisconsin. 1955. pp. 85. $1.15.

MAYO, H. B. Democracy and Marxism. New York: Oxford University Press. 1955. pp. 364. $4.00.

McHARGUE, DANIEL S. Michigan Government in Brief. Ann Arbor, Michigan: University of Michigan Press. 1955. pp. 66. $1.50.

MEISTER, IRENE W. Soviet Policy in Iran, 1917–1950: A Study in Techniques· Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 6. [Thesis Abstract, 1954, No. 6.]

MENDÈS-FRANCE, PIERRE AND ARDANT, GABRIEL. Economics and Action. New York: Columbia University Press. 1955. pp. 222. $3.50.

MOWBRAY, ALBERT H. AND BLANCHARD, RALPH H. Insurance: Its Theory and Practice in the United States. New York: McGraw-Hill Book Company, Inc. 1955. pp. 569. $6.00.

NATIONAL BUREAU OF ECONOMIC RESEARCH. Business Concentration and Price Policy. Princeton, New Jersey: Princeton University Press. 1955. pp. 514. $9.00.

NEWMAN, WILLIAM H. AND LOGAN, JAMES P. Management of Expanding Enterprises. New York: Columbia University Press. 1955. pp. 125. $2.75.

OLTON, ROY. Problems of American Foreign Relations in the African Area during the Nineteenth Century. Medford, Massachusetts: The Fletcher School of Law and Diplomacy. 1954. pp. 8. [Thesis Abstract, 1954, No. 7.]

PERROT, MARGUERITE. La Monnaie et l'Opinion Publique en France et en Angleterre 1924/1936. Paris: Librairie Armand Colin. 1955. pp. 252.

REILLY, GERARD D. States Rights and the Law of Labor Relations. Washington. D. C.: American Enterprise Association, Inc. 1955. pp. 35. $1.00. [No. 457 in the Series "National Economic Problems."]

SAMUELSON. PAUL A. Economics: An Introductory Analysis. New York: McGraw-Hill Book Company, Inc. 1955. pp. 753. $6.75. [Third Edition.]

SAMUELSON, PAUL A.; BISHOP, ROBERT L.; AND COLEMAN, JOHN R. Readings in Economics. New York: McGraw-Hill Book Company, Inc. 1955. pp. 488. $4.50. [Second Edition.]

SAXENA, S. K. Nationalization and Industrial Conflict: Example of British Coal⁻ Mining. The Hague, Netherlands: Martinus Nijhoff N. V. 1955. pp 185·

SHACKLE, G. L. S. Uncertainty in Economics and Other Reflections. New York: Cambridge University Press. 1955. pp. 267. $4.75.

SMITH, GEORGE ALBERT AND CHRISTENSEN, C. ROLAND. Policy Formulation and Administration: A Casebook of Top-Management Problems in Business. Homewood, Illinois: Richard D. Irwin, Inc. 1955. pp. 749. $6.00. [Revised Edition.]

SMITH, HOWARD R. Economic History of the United States. New York: The Ronald Press Company. 1955. pp. 763. $6.00.

SMITHIES, ARTHUR AND BUTTERS, J. KEITH. Readings in Fiscal Policy. Homewood, Illinois: Richard D. Irwin, Inc. 1955. pp. 596. $5.00.

SRIVASTAVA, K. N. Industrial Peace and Labour in India. Allahabad, India: Kitab Mahal. 1954. pp. 499.

STANKIEWICZ, W. J. AND MONTIAS, J. M. Institutional Changes in the Postwar Economy of Poland. New York: Mid-European Studies Center. 1955. pp. 125. $1.00.

STARK, W. Jeremy Bentham's Economic Writings. New York: Burt Franklin. 1955. pp. 600. [Volume III.]

SVENNILSON, INGVAR. Growth and Stagnation in the European Economy. New York: Columbia University Press. 1955. pp. 342. $4.50.

TAYLOR, O. H. Economics and Liberalism. Cambridge, Massachusetts: Harvard University Press. 1955. pp. 321. $5.00. [Harvard Economic Studies, Volume XCVI.]

UNITED NATIONS. Economic Survey of Europe, 1954. New York: Columbia University Press. 1955. pp. 313. $2.50.

———. Growth and Stagnation in the European Economy. New York: Columbia University Press. 1954. pp. 342. $4.50.

———. Handbook of Statistical Organization. New York: Columbia University Press. 1955. pp. 138. $1.50.

———. Methods of National Income Estimation. New York: Columbia University Press. 1955. pp. 58. 50 cents.

UNITED NATIONS ECONOMIC AND SOCIAL COUNCIL. International Bibliography of Economics. New York: Columbia University Press. 1955. pp. 429. $7.50.

UNITED NATIONS, FOOD AND AGRICULTURE ORGANIZATION. Use of Agricultural Surpluses to Finance Economic Development in Under-developed Countries. Rome, Italy: United Nations Food and Agriculture Organization. 1955. pp. 65. [Commodity Policy Studies No. 6.]

VARSHNEY, R. L. India's Foreign Trade During and After the Second World War. Allahabad, India: Kitab Mahal. 1954. pp. 340.

No. 277

THE
QUARTERLY
JOURNAL OF
ECONOMICS

FOUNDED 1886

PUBLISHED BY HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

The Quarterly Journal of Economics

Published by Harvard University

CONTENTS FOR AUGUST, 1955

THE
QUARTERLY JOURNAL OF ECONOMICS

FOUNDED 1886

CONTENTS FOR VOLUME LXIX

SUBJECTS

IN DEFENSE OF STATICS

By K. E. Boulding

I

It has for decades been a complaint against economic theory that it is static. The complaint was not raised much against the classical economists, probably because for the most part their theory was not static, but involved what Baumol has so felicitously called the "magnificent dynamics." With the coming of the marginal school and even of Marshall, however, economists lost interest in the great sweep of the "progress of society" and devoted themselves to equilibrium models of greater and greater refinement. The underworld voices of Marx and Veblen were raised in protest, but the main course of academic economics proceeded in majestic indifference to the dissidents.

Now, however, all is changed. It has become almost a matter of personal insult to call a theory "static," and even the most academic of economists is resolved to be dynamic at almost any price. In all this there is some gain. The time may not be inappropriate, however, to examine this dynamic enthusiasm, perhaps to utter a few cautionary remarks, and to reappraise the contributions of statics, especially in the general area of the usefulness of economics in guiding the making of decisions, both public and private.

For the clarification of the issues involved we are all much indebted to Samuelson.[1] He has made two important contributions.

1. Paul A. Samuelson, *Foundations of Economic Analysis* (Harvard, 1947).

The first is the clear distinction between comparative statics and process analysis. The second is the demonstration that the stability of an equilibrium system can only be established if the equilibrium system itself can be expressed as a limiting position of a dynamic model. In comparative statics we postulate an equilibrium system consisting, for instance, of a set of simultaneous equations with some sort of a solution. The equations contain parameters, which are regarded as the determinants of the system, and variables, which are regarded as determined by the parameters. The "solution" is that set (or sets) of values of the variables which is consistent with a given set of relationships and parameters as expressed in the equations. In comparative statics we compare one "solution" derived from one set of parameters with another solution derived from a different set of parameters. The great bulk of the analysis found in textbooks of economic theory is comparative statics — demand and supply analysis, the marginal analysis, and the Keynesian national income analysis all fall into this category.

By contrast, "true" dynamics or process analysis consists of a set of difference equations or differential[2] equations, also having a set of parameters, assumed to be constant, a set of variables, one of which is a "sequence variable" (for example, time) and a "solution" which in this case consists of a function giving the value of each variable at each point of "time." It is worth noting, however, that the concept of a sequence variable is more general than that of "time" in a calendar sense, and that it is perfectly possible to write process systems of difference or differential equations in which the sequence variable is some sort of space, or is any other variable, such as population, which follows a regular sequential course.

The importance of process systems lies in their use in *prediction*. There are two forms of prediction, both of which, however, depend on the existence of stable difference or differential equations. The first may be called "conditional prediction." It consists of a proposition of the form "if A then B in time t." It depends on the establishment of a stable difference equation or relationship relating A and B in a time sequence. Most of the propositions of science are of this

---

2. A differential equation is simply the limiting case of a difference equation when the standard "time" or sequential interval of the difference equation approaches zero. It is largely a matter of mathematical or computational convenience whether process systems are expressed as difference or as differential equations — there is no difference of principle. Some solutions are easier in one form, some in another. Numerical computations (for example, by electronic calculators) are usually performed by expressing the system in the form of difference equations. For "well behaved" systems there is usually little difficulty in expressing the system in either form.

form. Physics for instance does not, and cannot predict that there will be an atomic explosion on a given calendar date: it can, and does predict that if certain things are done at dates t_1, t_2, etc., there will be an atomic explosion at date t_n, where the intervals between t_1, $t_2 \ldots$ t_n are stable. The second form of prediction may be called "unconditional prediction." It is possible only where we have a closed system with perfectly stable parameters. So far only astronomy has had much success with unconditional prediction, and the success of the astronomers in predicting eclipses and less spectacular movements of the heavenly bodies has been the envy of their colleagues in other fields who have to struggle with less stable systems. It is not the virtue of the astronomers which leads to their success, however, but rather the virtue of the system with which they deal, which in simplicity and regularity exceeds any other scientific system. The geomorphologists and the geneticists deal with systems which have some characteristics of unconditional prediction, but with the advent of man unconditional prediction has become increasingly precarious even in these areas, and both land forms and mutations are in process of being profoundly modified by his activity. Indeed, it is an open question whether astronomy itself does not have to become a social science, for the continued rotation of the earth round the sun may come to depend on political rather than astronomical considerations.

For those who look to prediction as the supreme test of "success" in science, therefore, and especially for those who are enamored of unconditional prediction, the failures of the social sciences appear very glaring, and it is not surprising that they strain every nerve to make their systems "dynamic." There is a hunger for instance to make of economics an "astronomy of commodities" in which the various economic variables — prices, outputs, and so on — will dance in a lockstep as regular and precise as the music of the spheres, and in which the price of eggs is as predictable as the eclipses of the moon. I have no wish to deprecate any efforts to discover regular and predictable relationships in economic and social life. I shall argue, however, that the usefulness of such dynamic models as we have developed to date is extremely limited, and that while they throw some light on the nature of economic processes, they can be very misleading if taken too seriously. I shall also argue that the most secure propositions and the most reliable predictions, even though they are conditional predictions, arise out of comparative statics, and that when we are asked the awkward question "what good is economics to anyone," apart from its usefulness in providing a gainful occupation

for economists, the defense rests mainly on the achievements of rather old-fashioned comparative statics.

The great weakness of process models is that they are too easy to make and too rigid once made. Once we start to set up difference or differential equation models, there seems to be no end to the variety of models which can be constructed. It is tempting to think that all we have to do is to study the properties of these models until we find one which reflects in fair measure the properties of the real world. The search, however, is that for a needle in a haystack. Models can be made of varying degrees, with various lags or leads, with the many variables related in all kinds of different ways, all of which seem plausible, none of which stand out as more plausible than the other. Do we, for instance, want a model that exhibits a cycle? Practically all process systems involving difference or differential equations of the second order, and even some of the first order, exhibit cycles. The cycles can be made damped, exploding, or recurring by small modifications of the parameters or the relationships. Do we want a model exhibiting growth? Again a great variety of possible relationships gives us the required property, and there seems to be no great reason to select one rather than another.

At the other end, even if we have accomplished the task of selecting a model, it is almost certain to let us down empirically. Relationships and parameters which we must assume, for purposes of the model, to be constant turn out to be highly variable in practice. The trouble with predicting the course of history is that it seems to be a process system of almost infinite degree — rates of change of rates of change of rates of change ad infinitum are all relevant and determinant. Predicting the course of human events by means of a process model is rather like instructing someone to drive to a certain destination by means of a simple rule, such as "always take the first on the right and the second on the left, and repeat indefinitely." Such a rule may get the traveler somewhere, but it is most unlikely to get him where he wants to go, or to describe where he has been!

II

I propose to illustrate these strictures with five examples of process models which have significance for economics. The first of these is the simple model of exponential growth — that is, growth at a constant rate.[3] This is much beloved of the planners, both national and corporation. It is the basis, for instance, of the General

3. The basic difference equation for such a model is $x_{t+1} - x_t = Kx_t$, that is, the change in the variable in the sequence interval unit is proportional to the value of the variable. The solution of this model is $x_t = x_0 (k + 1)^t$. The corresponding differential equation is $\frac{dx}{dt} = kx$, the solution of which is $x = x_0 e^{\cdot\cdot}$.

Motors wage policy, and of most projections of national income. There is a magic 3 per cent, or perhaps 2 per cent, or something in between, at which everything in our society is supposed to grow — with occasional time out for wars and depressions. As a matter of fact, predictions based on this very simple model have been highly successful over moderately long periods. Many economic variables have a curious way of climbing back on the exponential bandwagon after a period of dissipation. Nevertheless it is a model that must be treated with extreme caution. One thing we are pretty sure of about growth is that with the possible exception of the whole expanding universe it cannot proceed at a constant rate forever. This is particularly true in a world where there are some principles of conservation, and in which therefore if one thing grows it is at the expense of something else. If the exponential model were sufficient, by this time there would only be One Thing in the universe, for anything which grew at a slightly faster rate than anything else would by now have gobbled up everything else. The whole biocene testifies to the impossibility of exponential growth, except for very short periods, and biologists have therefore devised more elaborate logistic models, in which absolute growth depends not merely on the size of the variable but also on the divergence of its present size from some equilibrium size. This model (in a great variety of special forms) gives us the familiar ogive or s-curve of growth. This, too, however, is likely to disappoint us, especially in the social realm. Growth processes in society get a new lease on life in the middle, or they collapse before they have worked themselves out. It seems sometimes as if parameters are most likely to change when we think them to be most stable, and that history is an elaborate practical joke with the predictors as butt.

III

The second dynamic system is that of population. The debacle of the population predictors is fresh in our memory. The system of prediction, however, is an elegant and quite sophisticated example of process analysis. The basic difference equations involved are survival or death functions, showing how many out of any given cohort of births survive into any given subsequent years. The simplest expression of such a relationship is a series of survival coefficients, s_1, s_2 ... s_i ..., the basic relationship being $a_t = s_i b_{t-i}$, where a_t is the total number in the population in year t of age i, and b_{t-i} is the total number of births in year $t - i$. In order to complete

the model there must also be a birth function of some kind relating the number of births in each year to some other variables of the model. This can be a simple function of time, extrapolating the past movement of births, or it can be as complex as we wish to make it. From these relatively simple materials the whole course of a population, its age composition and its birth and death structure can be predicted. The failure of recent predictions illustrates the treacherous nature of these dynamic models, even in cases where we thought we had good grounds for secure predictions. The one thing we do know about the future is that everyone who is alive today will either be dead or a year older this time next year! This knowledge is not enough, however, and it has been abundantly evident that the dynamics of population is much more complex and less predictable than we thought.

IV

Coming now to the third illustration, more specific to economics, we have the accelerator-multiplier models developed with such elegance by Samuelson.[4] These are dynamic Keynesian models. The simple multiplier model is based on two relationships: the composition of income, or savings-investment identity, Y_t (total income) $= A_t$ (accumulation or investment) $+ C_t$ (Consumption), and the consumption function, which expressed as a linear difference equation becomes

$$C_t = c + kY_{t-1}, \tag{1}$$

k being the "marginal propensity to consume" of this year's consumption related to last year's income. Combining these equations we get

$$Y_t = A_t + c + kY_{t-1}. \tag{2}$$

It is easy to show that this system follows a path of simple exponential decline of D_t, the difference between Y_t and the equilibrium value of Y_e,[5] assuming that A_t is constant from year to year. There is no

4. Paul A. Samuelson, "Interactions between the Multiplier Analysis and the Principle of Acceleration" (reprinted from the *Review of Economic Statistics*, 1939, in *Readings in Business Cycle Theory*, pp. 261–69).

5. The equilibrium value Y_e is obtained by putting $Y_e = Y_t = Y_{t-1}$ in equation (2), the equilibrium value being simply that which the process will repeat indefinitely. We have therefore, if A_t is constant and equal to A,

$$Y_e = \frac{A + c}{1 - k} \tag{i}$$

cyclical element in this model. A cyclical element is easily introduced, however, if we postulate an "accelerator" equation which relates A_t to the *change* in income,

$$A_t = a\,(Y_{t-1} - Y_{t-2}) \tag{3}$$

Equation (2) then becomes

$$Y_t = a(Y_{t-1} - Y_{t-2}) + c + kY_{t-1} = c + (a+k)\,Y_{t-1} - aY_{t-2} \tag{4}$$

This equation gives various types of solution, depending on the relative magnitudes of the parameters: it may give rise to damped, explosive, or stable oscillations: it may yield monotonic explosions away from or movements to equilibrium.

All this is very interesting, but does it really throw much light on the real forces of economic history? One may be permitted a few doubts. The doubts, of course, concern the stability of the parameters concerned. The famous failure of the postwar predictions of dire unemployment cannot fairly be imputed to intoxication with dynamic models, for this prediction was based essentially on the comparative statics, not the dynamics, of the Keynesian system. The breakdown of these predictions, however, arose from a deplorable (from the point of view of the predictors) or happy (from the point of view of society) upward instability in the consumption function — that is, essentially in the "*c*" parameter of our simple linear model. We all know now that the consumption function is far from being the law of the Medes and Persians which was sometimes hoped (or feared).

If even the stability of the consumption function is in doubt, there is double doubt about the stability of the accelerator function. There were shreds (pretty badly shredded by now, but still shreds) of evidence for the existence of a moderately stable consumption function. Nobody to my knowledge has ever put forward even a shred of evidence for a stable accelerator function. The accelerator is a pure construct of theoretical imagination, like Mrs. Harris. If we are honest we have to admit that we *have* no investment function, and that what is worse, we do not even know what we should put in it if we had one. Investment-planned or desired investment, that is, is the Holy Spirit of the Keynesian system. On it the whole fine edifice depends, but it insists on blowing where it listeth. In Keynes himself, of course, it depends on something which isn't in the system

$$D_t = Y_e - Y_t = \frac{A+c}{1-k} - (A+c) - kY_{t-1} = k\left(\frac{A+c}{1-k} - Y_{t-1}\right) = kD_{t-1} \tag{ii}$$

Equation (ii) as we have seen is simply the equation for exponential decline, as $k < 1$.

at all — a curious state of mind called the Marginal Efficiency of
Capital. But where this state of mind comes from, or goes to, is a
deep mystery which the system does not reveal. The accelerator is a
pebble cast hopefully into this chasmic gulf in the system. The most
one can say for it is that at least nobody ever attempted to predict
anything by it, and it remains a fine tour-de-force, retaining its hold
on the professional mind by its eminent suitability for regurgitation
in examinations.

V

The fourth dynamic system to be brought under review suffers
from defects like the third. This is the Harrod-Domar-Hicks dynamic,
or H-D-H for short. This has the peculiar attraction of being a very
gloomy dynamic, which in part accounts for its success, economics
having a strong masochistic urge to perpetuate its reputation as the
Dismal Science. It may conveniently be expounded graphically in a
version which is somewhat more general than that of its progenitors.

FIGURE I

Thus in Figure I we have the familiar diagram in which national
income is measured on the horizontal axis and its components on the
vertical axis. MM_1 is the consumption function, ON_1 the 45° line
representing the basic identity $Y = A + C$. With income equal to
OL_1 consumption will be L_1M_1 and "investment" (desired accumula-
tion) M_1N_1. Suppose now that OL_1 represents full employment, or

capacity output. As a result of the investment represented by M_1N_1 the capacity of the system will rise, say to OL_2. If full employment is to be maintained in the next period, then, investment will have to be equal to M_2N_2. This investment again raises capacity to OL_3, and in the next period investment must be M_3N_3 if full employment is to be maintained. The maintenance of full employment therefore requires the continual growth of investment, and it is not hard to show that if the propensity to consume is constant and the "growth factor," the increase in capacity which is caused by unit investment, is also constant, this growth of investment must be at a constant rate — that is, the growth must be exponential.[6] This rate of growth is Harrod's "warranted rate." In the present case it is shown in note 6 that the excess of income over the "no accumulation income," OL, also grows at the warranted rate. This can be shown geometrically by the properties of similar triangles:

$$\frac{M_1N_1}{KN_1} = \frac{M_2N_2}{KN_2} \text{ and } \frac{KN_1}{KK_1} = \frac{KN_2}{KK_2}: \text{ whence } \frac{M_1N_1}{KK_1} = \frac{M_2N_2}{KK_2}.$$

Harrod's model is a special case of the above in which it is assumed that the consumption function goes through the origin. In this case the no-accumulation income is zero, and income itself also must grow at the warranted rate of growth in order to maintain full employment. There seems to be no need, however, for this very restrictive assumption.

6. Let A_0, A_1 be the amounts of investment at times t, $t + 1$. Then
$A_1 = M_2N_2 = S_2R_2 - S_2M_2 + R_2N_2 = M_1N_1 - kM_1S_2 + M_1S_2 = A_0 + (1-k)rA_0$
where k is the marginal propensity to consume and r is the ratio $\dfrac{L_1L_2}{M_1N_1}$ or

$\dfrac{\text{increase in capacity}}{\text{amount of investment}}$, the growth factor. That is,

$$A_1 = A_0(1 + r(1 - k)) \tag{i}$$

Investment therefore grows at a constant rate, $r(1 - k)$, if r and k are constant. Suppose now the equation of the consumption function is $C = c + mY$. Then

$$Y_0 = C_0 + A_0 = c + mY_0 + A_0, \text{ or } Y_0 = \frac{c}{1 - m} + \frac{A_0}{1 - m} \tag{ii}$$

Now let y be the income at which accumulation is zero (OL in the figure). Then

$$y = c + my, \text{ or } y = \frac{c}{1 - m} : \text{ hence from (ii)}$$

$$Y_0 - y = \frac{A_0}{1 - m}$$

A_0 however grows exponentially at the rate $r(1 - k)$. $Y_0 - y$ therefore must grow at the same rate.

The dismal character of the model arises out of the further hypothesis that the warranted rate of growth which is necessary to maintain full employment is greater than the "natural" rate of growth which the growth of population and technology permit. If this is so, the only way to have full employment is to grow faster than is possible in the long run: the only way to do this is to grow from a position of unemployment, in which case growth of income at a faster rate than the long-run "natural" rate of growth is possible — but only as long as unemployed resources remain to be absorbed. Once growth hits the ceiling of full employment it must come down to the natural rate. But it cannot stay at the natural rate: it has to be above the natural rate to maintain full employment. Hence income must come down in a depression until it is low enough to start the process all over again. This is the Hicks, or Hiccup, theory of the trade cycle.

It cannot be denied, I think, that this model sheds at least one or two rays of light on the real world — mainly, however, through its defects rather than through its virtues. Its great defect is precisely its rigid dynamism — the fact that it assumes a number of parameters constant which are almost certainly never constant for long. Its gloom is derived almost entirely from the assumption that the consumption function is stable — an assumption which we have already seen to be constantly violated by plain fact. It is perfectly possible to chase away the gloom, therefore, by assuming a sufficient upward movement of the consumption function in response to increasing accumulation of capital. Thus in Figure I suppose that in the second period the consumption function rose to the dotted line $M'M_2'$. The amount of investment required to maintain full employment is now $M_2'N_2$ which may be actually less than M_1N_1. We could perfectly well postulate a dynamic system, therefore, in which the consumption function rose steadily to the point at which there was no further accumulation and the system reached a stationary state with full employment and consumption equal to income! This assumption is at least as inherently plausible as the assumptions of the H-D-H system. OM we may suppose represents "fixed consumption," that is, that part of consumption which is a function of the capital stock, not a function of income. This fixed consumption clearly depends on the size of the capital stock, including the human population. The more people, livestock, houses, machines, etc., that exist in a society the more resources must be devoted to their mere maintenance, as anyone who has graduated from the ownership of a small house to a large house knows all too well. The maintenance is not automatic,

of course, but it still represents consumption: if capital is not maintained this appears as disinvestment. It is not difficult to formulate conditions under which the investment required to maintain full employment will grow, be stable, or decline.[7]

This modification of the H-D-H system, though I flatter myself that it is an improvement, and at least has a slight potentiality of engendering a mathematically qualified cheerfulness, nevertheless still exhibits the fatal defects of any simple dynamic system, that it assumes constancy in parameters which are almost certainly never constant for long. Consequently its value in prediction would be small even if we could identify the values at the present time of the parameters concerned.

VI

The fifth and last example of a currently popular dynamic system or class of systems is the recent dynamization of Malthus into an embarrassing variety of systems of economic development. Drs. Leibenstein and Haavelmo[8] are the principal practitioners of this art,

7. Investment required to maintain full employment will grow, be stable or decline as

$$MM' \underset{>}{\overset{<}{=}} M_2N_2 - M_1N_1 \tag{i}$$

Let us define d as the marginal durability of new accumulations, $= \dfrac{M_1N_1}{MM'}$. The more durable the additions to the stock of capital, the less will be the increase in fixed consumption necessary to maintain them. Then:

$$MM' = \frac{M_1N_1}{d} = \frac{L_1L_2}{rd} \tag{ii}$$

$$M_2N_2 - M_1N_1 = S_2N_2 - S_2M_2 - S_2R_2 = R_2N_2 - S_2M_2 = L_1L_2(1-k) \tag{iii}$$

Inequality (i) therefore reduces to

$$\frac{1}{rd} \underset{>}{\overset{<}{=}} (1-k)$$

That is, investment to maintain full employment must grow, be constant, or decline as

$$rd(1-k) \underset{<}{\overset{>}{=}} 1.$$

The more durable the additions to capital, the more investment increases capacity, and the less the marginal propensity to consume, the more likely are we to run into the situation where the maintenance of full employment requires perpetual growth of investment.

8. Harvey Leibenstein, *A Theory of Economic-Demographic Development* (Princeton, 1954). T. Haavelmo, *A Study in the Theory of Economic Evolution* (Amsterdam, 1954).

the first using difference equations, the second differential equations. Again we may conveniently illustrate the general nature of the systems with a diagram, Figure II. Here we measure population on the horizontal, per capita income on the vertical axis. We postulate a "diminishing returns curve" R_0P_0 relating per capita income to population under a given resource situation. This is assumed to be linear for the sake of convenience. We suppose furthermore that there is
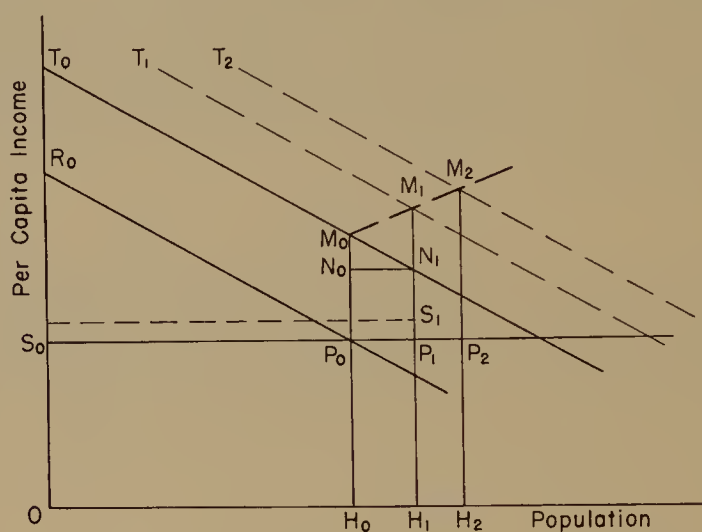
FIGURE II

some level of per capita income, OS_0, which is the "subsistence level," defined as that level of per capita income above which the population will grow, below which it will decline, and at which it will be stationary. The equilibrium level of population is S_0P_0, where the actual per capita income as given by the diminishing returns curve is at the subsistence level. Now suppose that there is an "improvement," reflected in a rise in the diminishing returns curve to T_0M_0. The first effect of this is to raise per capita income from H_0P_0 to H_0M_0. Now, however, there is a "surplus" — an excess of actual income above the subsistence level, P_0M_0. This "surplus" is the dynamic agent. It is likely to have two results: in the next period there will be an increase in population, to S_0P_1. There may also be an "induced improvement," raising the diminishing returns curve to T_1M_1. If there were no induced improvement the per capita income would fall in the first period as a result of the rise in population from H_0M_0 to H_1N_1. Because of the induced improvement, however, per capita income moves from H_0M_0 to H_1M_1 — which may be either a rise or a fall,

depending on whether the rise due to the induced improvement exceeds or fails to exceed the fall due to the rise in population. The "surplus" above subsistence is now P_1M_1, and if the relationships are constant we can project the model into the second year with population S_0P_2, per capita income H_2M_2, and surplus P_2M_2. The dynamic course of the population and the per capita income is then represented by the heavy dotted line $M_0M_1M_2$ — and can be projected as far as we wish. Probably the simplest system is that in which the ratio of population growth to surplus, $\dfrac{P_0P_1}{P_0M_0}$ is a constant, g, the ratio of induced improvement to surplus, N_1M_1/P_0M_0 is a constant, n, and the slope of the diminishing returns curve is a constant, d. It can then be shown that the "surplus" will either expand exponentially without limit, or will decline exponentially to zero again, at a rate of growth (or decline) equal to $n - dg$.[9]

A great variety of other models can be constructed, depending on the kind of relationship assumed between the variables; and depending on the model chosen, various time paths for the variables — that is, various properties of the model — will emerge. We can have models in which the per capita income first rises and then falls to equilibrium, we can have models in which the nature of the path depends on the size of the initial disturbance, and so on. A very useful modification might be to suppose that a surplus gave rise to a change in the subsistence level itself. Suppose for instance in the figure that the surplus P_0M_0 created a rise in the subsistence level itself from H_0P_0 to H_1S_1. Then the surplus in the second period would be S_1M_1, not P_1M_1, and the dynamic would be modified accordingly. No matter how we modify the model, however, in its difference equation or differential equation form it again suffers from the fatal defect of constancy in parameters. It is extremely unlikely that any parameters of a simple model such as I have described above, or even of much more complex models, will maintain constancy for very long. We actually know very little about the many subtle forces which affect population. We know still less about the forces which

9. We have $g = P_0P_1/P_0M_0$, $n = N_1M_1/P_0M_0$, $d = M_0N_0/P_0P_1$. Then:
$P_1M_1 = P_1N_1 + N_1M_1 = P_0M_0 - N_0M_0 + N_1M_1 = P_0M_0 (1 - dg + n)$.
That is:
$$\frac{P_1M_1}{P_0M_0} = (1 - dg + n).$$
The surplus, therefore, grows at a constant rate, $n - dg$. The bigger the induced improvement coefficient, n, and the smaller the population growth coefficient, g, and the diminishing returns coefficient, d, the more likely is the model to be explosive.

affect improvements. We do not even know very much about the forces of diminishing returns. It seems certain that only very loose relationships exist between any such vague quantity as a "surplus" and population, capital, and technological movements. It seems probable that whatever relationships occur exhibit "diminishing returns" or perhaps even "increasing returns" with time — forces spend themselves, or sometimes refuel themselves. An impetus to population may peter out: an impetus to technological change may reinforce itself. In any case the dynamics is far beyond the capacity of any simple system of difference or differential equations to portray.

VII

There are two possible avenues of escape from this fatal rigidity of the dynamic model. One is the use of stochastic variables, by which we add or subtract a number picked out of a hat to whatever answer our difference equations have given us in each period. This has some advantage in formalizing the ever present uncertainty, and in some cases even leads to fairly well determined equilibrium solutions, but it has the disadvantage from the standpoint of prediction that it yields usually a rapidly widening range of possible values rather than a specific prediction. Furthermore, in the absence of any real knowledge of probabilities it may be doubted whether the stochastic model does more than formalize ignorance in an elegant and perhaps even misleading manner.

The second possible avenue towards useful dynamic models is the development of short-range models with large numbers of variables in the manner of Lawrence Klein.[1] This is an avenue which is well worth exploring. It is a recognized principle of physical science that if a system does not predict it is because essential variables are missing from it. If we try to predict the volume of a gas from the pressure alone, variations in temperature will throw our prediction out severely. If temperature is included in the model very nice predictions can be obtained. If molecular size is included still better predictions follow, and so on. There seems to be no reason why the same principle should not apply in social systems, where the number of essential variables is much larger than in physical systems. All power, therefore, to those brave souls who penetrate the n-dimensional wilderness shouldering the heavy artillery of the electronic calculator,

---

1. Lawrence R. Klein, *Economic Fluctuations in the United States, 1921–1941* (Wiley, 1950).

especially if they are decently sceptical about their results, and are sensitive to the empirical frailty of empirical regularity. We primitive tribesmen who still cling to the bows and arrows of Demand and Supply will cheer you wistfully from the sidelines.

VIII

It is no deprecation of heavy artillery, however, to affirm that it is unfitted to the swatting of mosquitoes. I shall conclude, therefore, with a small paean of praise to comparative statics, as not only the mainstay of the textbook but the main justification for the usefulness of economics, on the grounds that there are a lot of mosquitoes which may be swatted with these primitive weapons. Suppose we ask ourselves the slightly embarrassing question *"To whom* is a knowledge of academic economics useful?"— apart, of course, from the academicians? I think we would have to confess that many business men get along admirably without more than a trace of economics. Even though we may, and do, comfort ourselves with the reflection that tennis players get along equally well without a knowledge of ballistics, it is slightly humbling to realize the magnificent irrelevance of most economic theory to the daily conduct of small affairs. As we move towards the large affairs, however, the maker of tennis balls may need to know something about ballistics, and the maker of large decisions — the central banker, the statesman, perhaps even the large corporation executive — may need to know something about economics. In making decisions about high economic policy we have a good deal of confidence that the special skills of the economist have relevance — not that they are sufficient, but that better decisions are likely to be made with them than without them. When we ask "what skills are relevant," however, I think we must admit that at least at present it is the skills of comparative statics and not our very rudimentary skills in dynamics which are crucial.

There are two broad areas in comparative statics which yield fruit for the policymaker. The first is the field of price theory, where anyone who ventures into the control or manipulation of prices certainly needs to know a little elementary demand-and-supply analysis. It is useful to know that in the case of any commodity there is usually some price above which we get into one kind of trouble which may be described roughly as "surplus" and below which we get into another kind of trouble described roughly as "shortage." Even this simple knowledge might have saved unwary statesmen from committing a

vast number of economic boners over the long course of history. It is useful to know that price control almost always involves some form either of production control or consumption control. It is useful to know something about elasticities of supply and demand if one is going to tax commodities. It is useful to have some notions of general equilibrium, even if it is only to prevent total surprise when the solution of one problem creates three others. It is perhaps even useful to know that the power of monopoly depends on the absence of substitutes. All these little pieces of knowledge are simple in the extreme, and yet a surprising amount of trouble has been caused by ignorance of them.

As one moves towards problems which essentially involve dynamics, however, even in price theory, the beautiful simplicity of elementary knowledge is lost, and the economist takes on more and more the role of the magician. The sad state of antitrust policy is a case in point, where the contribution of the economist, especially in the area of monopolistic competition, has been a quite spectacular befuddlement, with the Department of Justice riding off on white horses in at least three different directions at once. The confusion here originates mainly in the fact that the problem of monopoly in the real world contains so many dynamic elements which cannot be treated under the simple rubrics of comparative statics. In terms of economic development monopoly seems to be less villainous than it is in the two-dimensional world of comparative statics, and we seem to have no adequate criteria to tell us where virtue ends and villainy begins. The same difficulty applies to the problem of economic development in general. The self-confidence with which the economist advises a price administrator collapses into frantic calls for an anthropologist in the house when the economist is roped into advising the government of an underdeveloped area on how to get rich quick.

The same picture is revealed in the other great area of economic competence, which is national income policy. This is probably the most impressive, even though it is the most recent area for the exercise of economic skills. The Keynesian contribution is, however, again essentially comparative statics — it is underemployment equilibrium which is the central concept of the system. Yet an appreciation of the simple Keynesian model, elementary as it is, makes the difference between being completely baffled by the problem of unemployment and depression and understanding its essential nature. On the policy side it means the difference between striking out wildly in all directions, and having at least a qualitative notion of the kinds of

things that ought to be done. The statesman who knows that in a period of depression one should *not* at any rate increase taxes, make desperate efforts to balance the budget, and raise interest rates may not know much, but he perhaps knows most of what economics has to offer him, and that knowledge may make the difference between the survival or the nonsurvival of his society.

Again, however, when we come to the true dynamics of the problem the situation is confused and difficult. The problem of *when* to do things — how to interpret the economic signals — is a problem of true dynamics, and on this point the hand of the economist fumbles. We worry perhaps that ill-timed policies may aggravate fluctuations instead of alleviating them, but on the detailed matter of the when and the where it is still the intuition of the politician that governs and not the skill of the specialist.

IX

My conclusion, therefore, is that as economists we do not know very much, but we do know something that is not to be despised, and that what we do know is mostly comparative statics. This is not to deprecate the importance of economic dynamics: it is merely to register scepticism as to its existence. And our pretensions at dynamics in the shape of simple and rigid models of difference or differential equations must be seen for what they are — interesting exercises, and not much else. They have little value either for prediction or for policy. If dynamics is to be saved, it must be along two lines. On the theoretical side there is a real need for the development of a *qualitative dynamics* — process analysis which does not assume constancy in parameters, but which analyses the *general shape* of economic processes. Conceived in this sense there is something to be learned even from the accelerator models, from the post-Keynesian H-D-H dynamics, and from the dynamic Malthusian models of economic development. There is great need for the development of topological techniques in this field, analogous to the topological techniques which have been so useful in comparative statics, which enable us to describe the "shape" of processes without committing ourselves to detailed functions or constant parameters. It is to be feared that the enthusiasm for difference and differential equations on the part of our bright young dynamists has led them away from this more potentially fruitful area. On the quantitative side there is still hope for multi-variable econometric models. Even here, how-

ever, there is need for deeper theoretical work, especially at the level of integration of micro- and macroeconomic behavior. It is not enough merely to thresh around in the n-dimensional field, adding variables here and relationships there as fancy takes us. The growth of models should be in some sense "organic" and should be inspired by some principles of growth which are derived from fundamental theory. It is at this point that the long-awaited integration of psychology and economics may begin to take shape. We are looking for stable relationships among quantities differing in time position. What this involves on the theoretical side is learning theory, broadly conceived — the relationship between past experience and present behavior. What is learned here at the level of the individual cannot fail to affect our models of the aggregate, and may lead to that organic growth of testable model systems which is the dream of the scientist.

K. E. BOULDING.

UNIVERSITY OF MICHIGAN

RECENT BANK MERGERS*

By Charlotte P. Alhadeff and David A. Alhadeff

I. Introduction

Bank mergers in the United States absorbed 206 banks in 1954. This is the highest annual rate since 1934. The mean annual rate of mergers during the forties was only 81. The present accelerated rate of bank mergers began in 1952 (100 mergers) and increased somewhat in 1953 (115 mergers).[1] The sharp acceleration in the bank merger rate of the past few years, as compared with the previous decade, has increased apprehensions about the possible monopoly implications of increasing bank concentration. It is the purpose of this paper to examine the nature and extent of the recent mergers, to explore some of the reasons for these mergers, and to analyze some of the legal and economic implications of bank mergers, especially with reference to the competitive effects in banking markets.

The analysis of bank mergers during the late twenties and early thirties is not very useful in understanding the present merger movement. During the twenties, for example, bank mergers reflected the broader merger movement in the rest of the economy. A centripetal shift of population to urban centers induced still other mergers. Moreover, the trust business of banks grew rapidly during the twenties, and many mergers were consummated in order to establish banks quickly in the profitable trust business. By contrast, the merger movement of the thirties was basically defensive in character. The sharp increase in the rate of bank failures during that period stimulated bank mergers to forestall a still greater rate of bankruptcies.

---

1. The merger rate of the forties is computed from data in *Monetary Policy and the Management of the Public Debt*, Replies to Questions and Other Material for the Use of the Subcommittee on General Credit Control and Debt Management, Joint Committee on the Economic Report (1952), 82d Congress, 2d Session, Part I, p. 556. (Hereinafter, this document will be abbreviated to Patman Report.) The figures for 1952 and 1953 are from *Federal Reserve Bulletin*, Feb. 1953, p. 181, and Feb. 1954, p. 210.

II. Factors Behind the Recent Mergers

The analysis of bank mergers in this paper is based on 208 bank mergers which occurred between January 1953 and mid-1954.[2] This section analyzes the causes of mergers, and attempts to determine their relative importance in the current merger movement. In the absence of direct, detailed information about the banks participating in the present mergers, we have compiled the publicly available statistical information about these banks on a *bank-by-bank* basis. These data, together with more generally available information, make it possible to speculate on the plausibility of the various causes[3] which have been advanced to explain the present mergers, and, thus, indirectly, to evaluate their relative importance. There are two parties to a merger — the acquiring bank and the absorbed bank. In attempting to evaluate the relative importance of the various causes of mergers, we are primarily concerned with the initiating cause (or causes). The major initiating causes can originate either with the acquiring bank or the absorbed bank.

2. *Annual Reports* of the Comptroller of the Currency supply the names of merging banks for mergers involving national banks. A comprehensive list of all bank mergers is in "Discontinued Bank Titles," *Rand McNally Bankers' Directory*, final edition, 1954. This listing includes information about some mergers as late as July and August 1954. Although the Rand McNally figures are not "official," they are "secured from authoritative sources." However, a few mergers may have been omitted from this study because the information in Rand McNally is occasionally incomplete or ambiguous.

For each of the banks in this study, information on deposits and capital funds was obtained from *Rand McNally Bankers' Directory*, final edition, 1952, 1953, 1954. These figures are taken from bank balance sheets, as of June 31, of the respective years. In each case we selected the latest pre-merger figures available in Rand McNally. In a few mergers, it was not clear which bank had been absorbed and which was the acquiring bank. In these cases, we assumed that the larger bank was the acquiring bank, since this is usually the case.

Hereinafter, all statistical references to bank mergers in the period 1953–mid-1954, for which no source is given in a note, are based on our special study which utilized the foregoing data.

3. To weight definitively the causes behind the present mergers would require a detailed analysis of each merger. Circulating questionnaires among the participating banks might have provided some of this information, but this direct approach has serious limitations, too. First, when dealing with small numbers, the predictably partial response to such questionnaires would raise serious problems of statistical significance stemming from the sampling problem. Second, the questionnaire technique is handicapped by the well-known difficulties of extracting information about "real" reasons rather than "good" reasons from those questioned. Third, to distinguish real reasons from good reasons, operating ratios for the individual banks are essential, and these are simply not available to private individuals.

Management problems. Management problems[4] are often cited as a cause of bank mergers,[5] but opinion divides sharply on the weight assigned to management problems in causing bank mergers. A study of bank mergers in the Third Federal Reserve District concluded that "A conservative estimate might be that management problems have played a part in bringing about at least one-half of the mergers included in our study."[6] On the other hand, the Comptroller of the Currency has stated that the problem of aging managements with no suitable replacements available is "far from being the primary cause" of the recent mergers.[7] Management problems would appear to be sufficiently unambiguous to permit direct testing of their importance in causing mergers; in fact, they are not. More than one consideration is usually involved in a bank merger; certainly there is more than a single result. The critical question is whether management problems are a primary *cause* of mergers or merely *facilitate* them.

An indirect statistical test proved highly suggestive in attempting to answer this question. For the period 1953–mid-1954, each of the 208 mergers was classified by state. This test showed that 68 per cent of the mergers occurred in only 5 states,[8] 82 per cent in 10 states, and 90 per cent in only 15 out of the 48 states plus the District of Columbia. In 20 states, there was not a single merger during this period. This concentration of mergers would be meaningless if it reflected only the disparities in numbers of banks in the various states. In fact, however, although 68 per cent of the mergers took place in the top 5 states, only 17.4 per cent[9] of all the commercial banks were located in these states. Moreover, this concentration of

4. The most characteristic form of this problem is an aging management with no suitable replacements.

5. Professor Marcus Nadler, for example, has indicated that bankers must attract capable young men into banking; otherwise, the failure to do so "will accelerate the merger movement among banks." *New York Times*, June 4, 1952, p. 42.

6. Federal Reserve Bank of Philadelphia, *Business Review*, Jan. 1955, p. 6. Apparently this conclusion was based on conversations with bankers, for the Federal Reserve Bank of Philadelphia stated that " . . . we are largely relaying what we have been told by bankers participating in the branch and merger movement." *Ibid.*, p. 2.

7. Cf. Comptroller of the Currency, *Ninetieth Annual Report*, 1952, p. 4. It should be mentioned, however, that in his 1951 and 1953 *Annual Reports*, the Comptroller reported management problems as one of the six or four reasons, respectively, for mergers involving national banks. *Ibid.*, 1951, p. 2, and 1953, p. 9.

8. Pennsylvania, New York, Oregon, Ohio, and California.

9. Calculated for Dec. 31, 1953, from *Federal Reserve Bulletin*, May 1954, p. 536.

mergers by states is not peculiar to the period 1953–mid-1954. During 1947–1952, for example, the top 5 states had 49.9 per cent of the mergers.[1] Since management problems are alleged to be ubiquitous in the banking system, it is unlikely that their influence in causing mergers should be concentrated in a small number of states, and that two-fifths of the states should not have experienced a single merger during this period.[2] It is reasonable to conclude that although management problems may have been involved in the recent upsurge of bank mergers, they were not a major initiating factor.

Costs and profits. Higher operating costs and lower profits from operation have also been cited to explain the accelerated rate of bank mergers.[3] Presumably higher operating costs have impaired the earnings position of some banks, making them candidates for merger. In the absence of individual bank operating ratios, it is impossible to test this hypothesis statistically. It is possible, however, to test whether higher operating costs and lower profits are exerting pressure on the banking system in some systematic fashion, i.e., by size of bank.[4] If the absorbed banks are responding by merger to the economic pressures in their size group, the underlying reasons for the weakness of the size category would also explain the strength of the present merger movement. On the other hand, if banks are succumbing to mergers as marginal banks, irrespective of size, their marginal character probably stems from numerous causes, from which it would be difficult to develop any sort of systematic analysis or prediction about future mergers.

A size classification of the banks participating in the merger movement showed that about 56 per cent of the absorbed banks had

1. The 5 states are Pennsylvania, New York, New Jersey, Ohio, and California. This information is based on Rand McNally Bankers Directory, final edition, 1951, and final edition, 1953.

2. Management problems can effect the number of banks by leading to mergers or to voluntary liquidations which are not incident to succession, conversion, and absorption. There were only 13 voluntary liquidations of the latter kind for the period under study, 1953–mid-1954. Cf. *Federal Reserve Bulletin*, Feb. 1954, p. 210, and Aug. 1954, p. 903.

3. Cf., for example, Comptroller of Currency, *op. cit.*, 1951, p. 2, and Marcus Nadler, as reported in the *New York Times*, Feb. 6, 1952, p. 39.

4. Bank size and bank structure (whether branch or unit) are excellent guides in defining bank classifications along economically significant lines. Bank costs, earnings, and portfolio composition are significantly conditioned by bank size and bank structure. Cf. David A. Alhadeff, *Monopoly and Competition in Banking* (Berkeley and Los Angeles; University of California Press, 1954), chaps. V, VI, VII, and X. Since the United States is still predominantly a unit banking country, bank size is employed as the important classifying variable.

TABLE I

Size Distribution of Merging Banks, 1953–mid-1954

Size of Banks: Deposits (000 omitted)	Insured Commercial Banks (%)	Acquired Banks		Acquiring Banks	
		Number	% of Mergers	Number	% of Mergers
$1,000 and less......	13.7	25	12.0	5	2.4
$1,000–$2,000.......	22.6	38	18.3	4	1.9
$2,000–$5,000.......	33.2	53	25.5	17	8.2
$5,000–$10,000......	15.3	42	20.2	15	7.2
$10,000–$25,000.....	9.2	30	14.4	43	20.7
$25,000–$50,000.....	2.9	13	6.2	25	12.0
$50,000–$100,000....	1.4	2	1.0	21	10.1
More than $100,000..	1.7	5	2.4	78	37.5

Sources: For insured banks, see FDIC, *Annual Report*, 1953, p. 115. For merging banks, ee note 2, p. 504. Since some banks absorbed more than one bank during this period, there were not 208 individual acquiring banks. However, there were 208 individual banks acquired, and, hence, 208 cases involving an absorbing bank.

deposits of $5,000,000 or less (Table I). Only 2.4 per cent of the absorbed banks had deposits of $100,000,000 or more, and only 9.7 per cent had more than $25,000,000. Among the acquiring banks, 37.5 per cent had $100,000,000 and over, and 47.6 per cent had $50,000,000 and over. In short, the majority of absorbed banks were small banks, and they were acquired by large banks.

Table II compares the costs of different banks by size of bank. With 1950 as the base year,[5] i.e., on the assumption that cost and

TABLE II

Total Current Operating Expenses per $100 of Total Assets, 1953/1950 and 1953/1947, Insured Commercial Banks

Size of Banks: Deposits (000 omitted)	1953/1950 (%)	1953/1947 (%)
$1,000 and less....................	109.0	139.4
$1,000–$2,000.....................	111.2	140.3
$2,000–$5,000.....................	112.9	137.7
$5,000–$10,000....................	116.7	136.2
$10,000–$50,000...................	119.5	137.4
$50,000–$100,000..................	128.8	141.5
More than $100,000................	124.7	137.7

Sources: Computed from data in FDIC, *Annual Reports*, 1947, pp. 140–41; 1950, pp. 258–59; and 1953, pp. 114–15. Ratios are based on totals for each size category, and are not averages of individual bank ratios within each category. The FDIC figures were used because of their broad coverage of banks. However, Federal Reserve ratios (which are averages of individual bank ratios of member banks) give very similar results.

5. The test interval was determined as follows: 1953 was considered a suitable terminal year, because the merger rate sharply increased in that year compared with the mean rate of the forties. The determination of an appropriate base year presented more difficulties, because the selection of a base year carries implications about how long the pressures must exist before they culminate in a merger. It seemed to us that this question could not be categorically answered because of wide differences in the holding power of different banks. Accordingly,

profit pressures are translated into merger action fairly promptly, the cost relatives generally show an increasing progression by size of bank. With 1947 as the base year, i.e., on the assumption that cost and profit pressures build up over a period of years before they result in mergers, the figures show a roughly constant percentage increase in costs in all size categories, with a mean increase of 38.6 per cent and a deviation of only three percentage points on either side of the mean. The most that can be said for the cost hypothesis alone is that cost pressures were not greater in the categories with most absorptions than in any other categories; at worst, the cost pressures were least in precisely the size categories with the greatest concentration of absorbed banks. Moreover, the cost relatives increased most on a 1950 base and were matched by the others on a 1947 base in the two largest size categories with 47.6 per cent of the acquiring banks. By itself, the cost hypothesis is not related to any systematic pressure on the banking system.

Clearly, however, the costs hypothesis should also be examined in the broader context of its relation to profits. On this basis, the cost-profits hypothesis appears to gain validity. Table III shows the net profits after income taxes[6] for all insured banks by size of bank.

TABLE III

NET PROFITS AFTER INCOME TAXES PER $100 OF TOTAL CAPITAL ACCOUNTS, 1953/1950 AND 1953/1947, INSURED COMMERCIAL BANKS

Size of Banks: Deposits (000 omitted)	1953/1950 (%)	1953/1947 (%)
$1,000 and less	83.8	77.4
$1,000–$2,000	84.2	75.8
$2,000–$5,000	84.5	77.7
$5,000–$10,000	87.0	85.3
$10,000–$50,000	89.5	94.7
$50,000–$100,000	94.1	107.7
More than $100,000	99.1	105.4

Sources: Same as for Table II.

Except for two size categories in the 1953/1947 comparison, profits after taxes declined, compared with either 1947 or 1950. Irrespective of the base year, profits declined most sharply in size categories with large percentages of acquired banks. Moreover, almost half the

the hypothesis was tested for a short interval as well as for a longer interval. It should be mentioned that the ratios are averages for each group of banks. No information is available on the dispersion.

6. Net profit after taxes is the most relevant profit figure to consider in a merger decision based on low earnings.

acquiring banks were in the two most profitable size groups, i.e., profits decreased least or actually increased.

To confirm the cost-profits hypothesis, the profits pattern must also be consistent with the hypothesis after adjusting for the particular distribution of banks among the different size categories. On this last test, the cost-profits hypothesis conspicuously fails. Table I showed that 56 per cent of all absorbed banks in 1953–mid-1954 had $5,000,000 deposits or less, but this category comprised almost 70 per cent of all insured banks in 1953. Even on a random distribution of mergers among banks of all sizes, more than 56 per cent of the absorbed banks would be concentrated in the size category of less than $5,000,000 deposits. Similarly, only 27.4 per cent of all insured banks were in the $5,000,000 to $50,000,000 size categories, but these categories accounted for 40.9 per cent of all absorbed banks in the period 1953–mid-1954. Thus, relatively to the size distribution of banks, absorbed banks were concentrated disproportionately in these mid-categories. For the costs-profits hypothesis to hold, profits would have had to fall most in these categories, instead of falling to an intermediate degree as they actually did.

Higher operating costs and lower earnings have not exerted a *systematic* pressure upon banks to merge. Whether high costs and low earnings were responsible in important degree for the actual mergers is moot in the absence of individual bank operating ratios. Even if costs and profits pressures were important in the actual mergers, the merged banks succumbed because of individual weaknesses, and not as representatives of their economic class.

Loan limits — the 10 per cent rule. Analyses of earlier consolidation movements among banks ascribed an important role to lending restrictions as a motive for mergers. This point has been described by Steiner & Shapiro:

. . . bank mergers have been consummated in order to increase a bank's capitalization and deposits, thereby enabling the institution to furnish more adequate service to its customers. The growth in the size of business enterprises increased the credit needs of these firms. Since the maximum unsecured loan to one borrower is limited to 10 per cent of a bank's capital and surplus, an increase in the bank's capital base permits a larger maximum loan limit. The increased resources of the merged bank enable it to grant the larger loans made permissible by the increased capitalization.[7]

Similarly, with reference to the current merger movement, it has been stated that " . . . another important reason behind the rash of

7. W. H. Steiner and Eli Shapiro, *Money and Banking* (3d ed., New York: Henry Holt and Company, 1953), pp. 105–6.

bank marriages is the need to keep up with bigger demands for credit from expanding industry."[8] Clearly, the increased lending limits which often accompany a merger must be listed as an advantage in weighing the merits of a proposed merger. But that is not the question at issue.[9] What matters is whether the 10 per cent loan limitation is a major motive for the recent upsurge in bank mergers.

During the forties, gross national product more than doubled, and the size of business firms also increased substantially. However, bank growth was also substantial. Between 1941 and 1952, commercial bank deposits more than doubled, while the total number of commercial banks declined.[1] Although the growth rates of different banks were not uniform, bank growth generally was not stunted while business firms were growing. Thus, the mean size of commercial banks in 1941 was $4.9 millions, as compared with $12.3 millions in 1952.[2] Moreover, although the capital funds of most banks did not keep pace with the growth in deposits, the average increase in bank capital (which is directly pertinent to the issue in question) was very impressive. In 1941, total capital of commercial banks was $7,173,000,000, and, in 1952, $12,888,000,000.[3] Although over-all growth of banks in the decade preceding the recent upsurge in bank mergers was substantial, numerous individual banks might have felt the 10 per cent pinch. It was shown above that, in rough terms at least, big banks have been absorbing small banks in the recent mergers. It would follow from the 10 per cent hypothesis that the large banks (and, indeed, especially the largest size category which contained 37.5 per cent of the absorbing banks) felt most acutely the pressure of lending limits. This would seem to support those who argue that banks have been most restricted by lending limits in competing for big customers with the life insurance companies. Although commercial banks are hampered by the 10 per cent rule, their competitors are not.[4]

8. George Cruikshank, *Wall Street Journal*, Mar. 10, 1955. According to *Business Week*, this is also the view of J. Stewart Baker, Chairman of the Board, Bank of the Manhattan Company. "His view is simply that banks have to keep pace with development of the U. S. economy. If industrial agglomerations of capital get bigger, the banks serving them must do the same. So far, he says, the banks have lagged in size of units." *Business Week*, Feb. 12, 1955, p. 126.

9. This is not to deny that, for individual cases, the increase in limit is precisely what counts.

1. Cf. *Federal Reserve Bulletin*, May 1954, p. 477.

2. Computed from *ibid*.

3. *Ibid*.

4. Cf. *Business Week*, Feb. 12, 1955, p. 118.

The major loan competition between commercial banks and
insurance companies is for term loans. However, since banks have a
comparative advantage over insurance companies in any form of
lending as against investing, term loans are mostly negotiated by
commercial banks. Indeed, term lending by insurance companies is
often complementary rather than competitive with commercial banks.
Term loans have been syndicated between banks and insurance com-
panies, particularly when the maturity of the loan was greater than
that desired by the banks. The insurance companies then take the
longer part of the maturity, and the banks, the shorter.[5] It is
unlikely that the remaining area of genuine competition between
banks and insurance companies, for the limited part of the banking
system which is involved, is important enough to explain the sharp
increase in over-all rate of bank mergers during 1953–1954.

This presumption is bolstered by the nature of recent bank
acquisitions. In 50 per cent of the mergers studied, the absorbed
banks' deposits were 10 per cent or less than the deposits of the
acquiring bank. In 65 per cent of the mergers, the ratio was 20
per cent or less. Thus, large banks were acquiring banks significantly
smaller than themselves. Furthermore, numerous acquiring banks
did not limit themselves to one acquisition during this period. About
50 per cent of the acquired banks were taken over by only 37 acquir-
ing banks. Eighty-three banks were acquired by 36 "repeaters," and
56 banks by 28 "repeaters." In only about 15 per cent of the mergers
were the acquired banks not converted into branches. The over-all
picture is of large banks acquiring banks significantly smaller than
themselves, with numerous acquiring banks involved in more than
one merger, and conversion of acquired banks into branches. But
this is certainly a roundabout, cumbersome, and expensive way to
ease the 10 per cent pinch. It would be more efficient and certainly
more effective to merge two large banks directly, with a correspond-
ingly sizable increase in capital. Since the pressure of lending limits
is ostensibly great on both, both parties would be impelled to seek
each other out, and the merger would be mutually beneficial.

In this study, the expression "merger" covers mergers, absorp-
tions, consolidations and purchases.[6] For most purposes, this com-
prehensive use of the term is unobjectionable, but these different

5. Cf. Roland I. Robinson, *The Management of Bank Funds* (New York:
McGraw-Hill Company, 1951), p. 193.
6. Cf. Federal Reserve Bank of Philadelphia, *Business Review*, Nov. 1954,
pp. 10–11, for usual distinctions.

techniques can affect the resultant capital structures. For example, a bank purchase can increase[7] or decrease a bank's capital depending upon the particular terms of the merger. Among 170 mergers involving national banks between August 17, 1950, and December 31, 1952, 96 involved the purchase of assets and assumption of liabilities.[8]

Because of the foregoing, the limitations on bank lending imposed by the 10 per cent rule are not a persuasive explanation for the sudden acceleration in the over-all rate of mergers in 1953 and 1954. Again, this is not to deny that lending restrictions were important in individual cases. For the majority of participating banks, however, the available data do not support the 10 per cent hypothesis as a major initiating factor in the upswing of mergers.

Uneven rates of growth. This section proposes and tests the hypothesis that uneven rates of growth of different banks are a major reason for the increase in bank mergers. It was mentioned above that the banking system as a whole expanded tremendously during the period 1941 through 1952. During this period, for example, deposits increased more than 240 per cent.[9] This terrific rate of growth did not proceed smoothly and uniformly for all banks in the system. In the economy as a whole, during the expansion phase of the business cycle, different growth (expansion) rates of different sectors, or even of different firms within the same sector, create structural maladjustments which contribute to the instability of the upper phases of the boom. Similarly in banking, the very rapid expansion of the past decade has created stresses and maladjustments owing to uneven rates of growth of different banks. These maladjustments have been particularly reflected in bank capital ratios. After a prolonged period of rapid but uneven growth, some banks are left outside the bounds of a desirable capital ratio.[1]

The capital-deposit ratio has important implications both for the soundness and the profitability of a bank. When this ratio falls below the acceptable norm, because of a period of rapid growth, the safety of the bank is impaired by conventional standards. On the

---

7. Bank capital may be increased by purchase of a bank if new capital is raised by the acquiring bank in anticipation of (or shortly after) the bank purchase.

8. Cf. Comptroller of Currency, *op. cit.*, 1952, p. 5.

9. Cf. *Federal Reserve Bulletin*, May 1954, p. 477.

1. Over the years, the criterion of an acceptable capital-deposit ratio has varied, and properly so, for capital ratios have no special significance without reference to other variables, such as the general trend of business conditions, portfolio composition, etc. At any moment of time, however, there is general agreement on acceptable capital-deposit ratios.

other hand, a lagging growth rate may raise that ratio (and, *ceteris paribus*, conventional notions of soundness) but impair the earnings position of the bank. The banker's most difficult responsibility is to steer a course between the two extreme positions.

An examination of the capital-deposit ratio of the acquiring banks during 1953–mid-1954 suggests that the failure to achieve this intermediate position is an important cause of mergers by banks with above average capital ratios, but not for banks with below average capital ratios. Table IV compares the capital-deposit ratios

TABLE IV

Per Cent of Total Capital Accounts to Total Deposits

Size of Banks: Deposits (000 omitted)	FDIC Banks 1953	Member Banks 1953	Acquiring Banks 1953–mid-1954
Under $1,000	11.2	12.3	13.3[1]
$1,000–$2,000	9.6	10.1	14.5[1]
$2,000–$5,000	8.4	8.8	9.9
$5,000–$10,000	7.6	7.8	10.2
$10,000–$25,000	7.1	7.4	9.9
$25,000–$50,000	6.8	6.8	9.1
$50,000–$100,000	6.9	6.5	8.6
More than $100,000	7.6	7.1	8.2

Sources: For FDIC banks, ratios were computed from FDIC, *Annual Report*, 1953, pp. 114-15. For member banks, ratios were taken from Federal Reserve Bulletin, Aug. 1954, p. 901. For ratios of acquiring banks, see note 2, p. 504. Capital, surplus, and undivided profits were added to secure total capital accounts of acquiring banks.

The FDIC data are averages of aggregates for each size category, while the Federal Reserve figures are averages of individual bank ratios. (The ratios for acquiring banks are also averages of individual bank ratios.)

[1] There were only five or less banks in these categories. Cf. Table 1.

of the acquiring banks with the mean ratios for their particular size classes. It is clear from the table that the means of the absorbing banks' capital-deposit ratios were higher, and sometimes much higher, than the means for their size classes. It should be mentioned that the capital-deposit ratios of absorbing banks varied even within a single size category. On the average, however, the capital-deposit ratios of the acquiring banks were clearly higher than the averages for the banks in their size group.

The contention that banks with below average capital-deposit ratios do not seek mergers to correct their capital position is further confirmed by Table V, which shows the capital-deposit ratio of the absorbed bank as a ratio of the capital-deposit ratio of the absorbing bank. If banks with below average capital-deposit ratios were motivated to merge to correct their capital position, we should expect to find the acquiring banks merging with banks which had higher

TABLE V

RATIOS OF TOTAL CAPITAL/DEPOSITS OF ABSORBED BANK
TO TOTAL CAPITAL/DEPOSITS OF ACQUIRING BANK,
FOR MERGING BANKS, 1953–MID-1954

Ratio (%)	Number of Mergers	Ratio (%)	Number of Mergers
30.0– 39.9	3	130.0–139.9	10
40.0– 49.9	4	140.0–149.9	9
50.0– 59.9	5	150.0–159.9	7
60.0– 69.9	8	160.0–169.9	7
70.0– 79.9	22	170.0–179.9	5
80.0– 89.9	20	180.0–189.9	3
90.0– 99.9	25	190.0–199.9	0
100.0–109.9	28	200.0–299.9	14
110.0–119.9	15	300.0 and more	4
120.0–129.9	18		

Sources: See note 2, p. 504. Total capital is the sum of capital, surplus, and undivided profits. One case was omitted because of incomplete data.

capital ratios than the acquiring banks. However, the table reveals a clear clustering about the 100 per cent ratio, trailing off from this clear mode in either direction. Overwhelmingly, the acquiring banks absorbed banks with the same or very similar capital-deposit ratios. There is no clustering in the high ratios above 100 per cent.

These results are entirely reasonable. It seems unlikely that fast-growing banks with below average capital-deposit ratios would acquire other banks to improve their capital position. First, a bank with below average capital-deposit ratio could improve its capital position in a more direct manner than bank merger. Even with the same earnings on assets as other banks, the low capital-deposit ratios of banks with below average capital ratios result in higher earnings on capital, and permit them, *ceteris paribus*, to pay higher dividends. These higher earnings rates and pay-out rates would, in turn, facilitate the sale of more stock at favorable terms. Second, again on the assumption that banks with below average capital-deposit ratios are at least equally profitable on assets as other banks in their size class, the higher earnings on capital would permit them to increase capital directly by increasing surplus, i.e., by retained earnings, or alternatively, by paying stock dividends. Finally, many bank acquisitions are by purchase for cash of assets and assumption of liabilities, and purchases reduce capital-deposit ratios.

At first glance, it might appear that the distribution of Table V is also inconsistent with the capital hypothesis for banks with above average capital-deposit ratios. That is, it might be argued that such

banks should acquire banks with lower capital-deposit ratios — and the figures show no such pattern of acquisitions. However, to suggest that banks with above average capital-deposit ratios should absorb banks with lower capital-deposit ratios to correct the capital-deposit position of the former is to miss the main problem of such banks. Basically, it is presumed that the high capital ratios of such banks are owing to their lagging rates of growth relatively to banks in general. This lagging growth rate results in a high capital-deposit ratio, but the main problem of these banks is not the narrow technical problem of adjusting capital (as it is for banks with below average capital-deposit ratios). If the matter were merely technical, capital could be quickly reduced, and the capital-deposit ratio correspondingly improved, by increasing dividends rather than surplus. The basic problem of these banks is a growth problem, and a high capital-deposit ratio is merely a manifestation of that growth problem. The quickest way for banks with lagging growth rates to recoup (or simply to improve) their position is to acquire other banks, and thereby to increase their assets, and, hence, their earnings. Banks with above average capital-deposit ratios are banks which need more business. They can get more business quickly by acquiring other banks, and do not have to worry unduly about the relationship of the capital-deposit ratio of the acquired bank to their own.

Desire for branch bank expansion. The growing importance of branch banking during the last twenty years is producing a striking change in the formal structure of the commercial banking system in the United States. On December 31, 1934, branch offices comprised only 16 per cent of total commercial banking offices; by December 31, 1954, branch offices were 31 per cent of the total.[2] In two decades, the number of branch offices more than doubled, and branch offices virtually doubled as a percentage of total commercial banking offices.[3]

a. Reasons for growth of branch banking. The acceleration of the trend towards branch banking during the past two or three years is due to several factors. First, many branch banks have exhibited a spectacular rate of growth. Indeed, the largest bank in the country, the Bank of America, with resources exceeding nine billion dollars, has outgrown the traditional giants of the banking system, the Wall Street Banks. It is not necessary to review here the relative advantages or disadvantages of branch and unit banking. It is sufficient

2. Computed from *Federal Reserve Bulletin*, Feb. 1955, p. 208.
3. *Ibid.* There were 3,007 branches of commercial banks on December 31, 1934, and 6,306 branches on December 31, 1954.

for our purposes merely to note the rapid growth of branch banking and to suggest that the point of this experience has not been missed by many unit banks and smaller branch banks.

Second, branch banking helps to ease the earnings problems posed by the population shift from metropolitan centers to suburban and outlying areas. This movement has affected bank earnings in two ways. Population shifts have drained deposits from banking offices in formerly populous areas. Other things being equal, this loss of deposits raises the capital-deposit ratio, and reduces the leverage on bank earnings caused by a low capital-deposit ratio. Moreover, the shift of population has affected bank earnings directly, by transferring the loan business of depositors who have moved to local banks in the new domicile. This loss of loans is particularly serious, because of the volume which loans to individuals has attained during the last few years. Loans to individuals are of two major kinds: consumer loans and real estate loans for home purchases and improvements. Both categories have increased sharply in recent years, not only in absolute terms, but as a percentage of total loan portfolios of commercial banks. For example, on December 31, 1947 consumer loans were $5,723,000,000, or 15.0 per cent of total commercial bank loan portfolios. By June 30, 1954, these loans amounted to $14,462,000,000, or 21.5 per cent of total loan portfolios. Similarly, consumer loans and real estate loans were 39.8 per cent of total loans on December 31, 1947, and had increased to 47.1 per cent of total loans by June 30, 1954.[4]

The increasing importance of the consumer loan is an integral part of a broader phenomenon — the growing importance of retail banking and the relative decline of wholesale banking. Because of the "... impact of steeply graduated income taxes on the wealthy and the growth at the same time of middle-income groups ... banks have sought to replace a smaller number of large accounts with a much larger number of accounts drawn from the middle-income group."[5] Branch banking is the obvious answer for wholesale banks desiring to become more deeply involved in retail banking.

Noneconomic factors have also stimulated the trend toward branch banking. Considerations of prestige, or the desire for sheer size of bank are no doubt important in stimulating individual bank

4. The last two figures overstate the amount of loans to individuals, because the real estate loans include all such loans and not merely those for home purchases. All figures are from the *Federal Reserve Bulletin*, Feb. 1955, p. 162.

5. Chase National Bank (Address by David Rockefeller, Senior Vice President, March 11, 1955), "Banking's New Look," p. 10.

expansion — and an excellent way to become a banking giant is to expand the number of branches.

  b. Acquisition vs. de novo branches. The drive to establish and expand branch banking systems has proceeded along two main lines: establishing *de novo* branches, and acquiring existing banks and converting them into branches of the acquiring bank. In 1942, for example, only 28 *de novo* branches were opened by commercial banks. In 1948, this figure had risen to 151; and in 1954, 341 *de novo* branches were opened. Branch banking has also expanded by bank mergers. In 1942 and 1943, slightly less than one-third of the banks acquired in all commercial bank mergers became branches of the acquiring banks. By 1948, this figure had risen to 75 per cent, and, by 1954, 85 per cent of all acquired banks were converted into branches of the acquiring bank.[6] Acquisitions rather than *de novo* branches are often motivated by both business and legal reasons. When a bank is acquired, and converted into a branch, the staffing problem is usually solved automatically by retaining most of the old personnel. Far more important, the branch does not have to build its deposits from zero, and it benefits from existing banker-customer loan relationships. The absorbing bank acquires not only the assets and liabilities of the merged bank; it acquires a going institution. Moreover, the acquiring bank's share of the local market is prospectively larger than it would be by establishing a *de novo* branch in the same community without at the same time replacing any of the existing banks. Legal considerations also often impel branch bank expansion by acquisition rather than by *de novo* branches. A *de novo* branch must satisfy the regulatory authorities of the "need" for its existence, and that is not always easy when other banks already exist in the area. Moreover, it would be difficult to obtain authorization for as many *de novo* branches as could be obtained in a large, multi-bank merger. Indeed, some state laws make it easier to acquire a bank and convert it into a branch than to open a *de novo* branch.[7]

  c. Merger clusters. Given this basic desire to develop and expand branch bank systems, a merger movement can get under way by the

  6. For 1942, 1943, and 1948, Patman Report, *op. cit.*, Part I, p. 556; for 1954, *Federal Reserve Bulletin*, Feb. 1955, p. 208.
  7. Cf. Board of Governors, Federal Reserve System, *Compilation of Federal and State Laws Relating to Branch Banking Within the United States* (July 1, 1951). Hereinafter, this will be cited as *Compilation*.
  These reasons explain why acquisitions are generally preferable to *de novo* branches, given a choice between the two. It should be stated, however, that more branches are opened *de novo* than by acquisition and conversion of other banks.

aggressive action of a comparatively small number of banks.   To be sure, aggressive action by a small number of banks may not be suffi- cient to *constitute* a merger movement, owing to the large size of the banking system, the prohibitions against interstate branch banking, and the limited character of branch banking in some states.   However, the aggressive acquisitions by a few banks are often sufficient to *generate* a wave of mergers in the banking system.   This clustering of mergers is due to two factors.   Banks which have not yet begun to acquire other banks begin to do so in simple imitation of the initiators. A pattern of acquisitions which is desirable for the initiators can be equally desirable for the imitators.   The clustering is also explained by the defensive reaction of nonparticipating banks to the acquisition activities of the initiating banks.   Once mergers have begun in an area, the relative standing of the other banks is either actually or imminently threatened,[8] and a retaliatory merger is the surest way to maintain a threatened position.[9]   The exact timing of a merger movement hinges on the determination by a few aggressive banks that the over-all situation is "favorable."   Apparently that deter- mination had begun to crystallize by 1952, and in 1953 and 1954 a cluster of mergers began to develop.

d. Initiating cause vs. facilitating condition.   The two major initiating causes of bank mergers identified above pertain to the motives of the acquiring bank.   Some of the factors alleged to be major initiating causes of recent bank mergers are more properly viewed as facilitating conditions, for they explain why the absorbed bank was willing to be acquired by another bank.   Management problems, discussed earlier, generally fall in this latter category. Similarly, the willingness of banks to be absorbed may be enhanced by the fact that banks are often worth more dead than alive.[1]   This is

---

8. Although it was denied by a spokesman for the National City Bank, it is widely reported that the merger of the National City Bank with the First National Bank was, at least in part, motivated by National City's reluctance " . . . to play second fiddle to Chase Manhattan for long after being ahead of Chase in deposits through most of the postwar period." *Business Week*, Mar. 5, 1955, p. 32.

9. This interpretation is also consistent with the observed geographical clustering of mergers by particular states.

1. For specific examples, cf. *New York Times*, May 25, 1952, p. F6, and *Business Week*, Aug. 26, 1950, p. 69.   There are many reasons why the stock- holders of a merged bank might realize more upon merger than the market value of the stock.   First, published statements of banks often understate the actual value of bank buildings.   Second, conservative bank management typically pre- scribes the carrying of marketable assets at their market or cost value, whichever is lower.   This practice understates the true value of such assets during periods

commonly held to be "the primary cause" of bank mergers,[2] but it is more properly viewed as a facilitating condition rather than as an initiating cause of mergers. Surely it is difficult to conceive of bank management actively seeking the demise of its institution simply because the corpse value of the bank is greater than its living value. However, an attractive price offered by an eager buyer might induce even a sound bank to give up its corporate existence. This is further suggested by the observed geographical concentration of mergers.[3] The tendency for book value of banks to be underrated by the market value of bank stocks is a general condition of the banking system during certain periods. If this were, indeed, "the primary cause" of mergers, the merger movement would not be highly concentrated geographically.

III. Government Regulation of Bank Mergers

An intricate web of legislation envelops most banking activities, including bank mergers. Bank merger laws seek to protect stockholders, depositors, borrowers, and the public interest. Specific bank merger legislation sets forth the conditions for approval of the merger by the chartering or insuring agency. The conversion (or transfer) of branches in banking mergers is regulated by branch banking legislation. Bank mergers are also affected by state and federal laws for the general regulation of competition and monopoly.[4]

Specific bank merger legislation. Much specific bank merger legislation of the last forty years has made it easier for banks to merge[5] within and across jurisdictional chartering boundaries. In

of rising security values. Third, although good will is often carried at a nominal sum, its true value may be realized in a merger. Finally, low bank dividends compared with industrial earnings may hold market value below book value.

2. For example, in his 1952 *Annual Report*, the Comptroller of the Currency noted the current trend of mergers and stated that "The primary cause rests in very attractive purchase prices which the shareholders of the solicited banks cannot see their way clear to refuse." Cf. Comptroller of Currency, *op. cit.*, 1952, p. 4.

3. A special check of our figures showed that the geographical concentration of mergers in five states is very high even when compared with the ratio of banks in these five states to banks in all states permitting branch banking. Hence, permissive branch banking legislation is not a sufficient explanation for the geographical concentration of mergers.

4. In the following discussion, each type of legislation is treated briefly. The interested reader can find more details and the exact wording of particular legislation in the sources mentioned in the notes.

5. In the discussion of specific bank merger legislation, "purchase" mergers are not included under the terms merger or consolidation, unless specifically so noted.

1918 an amendment to the National Banking Act (the Consolidation Act) made it possible for one national bank to consolidate directly with another national bank. In 1927 the McFadden-Pepper Act provided for the direct consolidation of national and state banks under the national bank charter.[6] In 1950 the "two way street" Act permitted national and state banks to consolidate under the charter[7] of either authority.[8]

Because of the dual banking system, 49 chartering jurisdictions control bank mergers, but the restrictions on mergers are primarily technical. These technical provisions[9] describe (1) procedures for obtaining approval of a majority of the Board of Directors and of owners of two-thirds of the capital stock of the affected banks, (2) the steps to be taken by stockholders of either bank who dissent from the terms of the merger,[1] and (3) the effects of the merger upon the rights and obligations of the merging institutions. Banks desiring to merge must submit detailed plans for the approval of the appropriate banking authorities.

Other technical hurdles beset certain mergers (including purchase mergers) of insured banks which reduce the combined capital and surplus of the merging institutions.[2] Moreover, regardless of the effect on bank capital, all mergers of an insured bank with (or a transfer of assets to, or an assumption of liabilities by) an uninsured bank must secure the approval of the Federal Deposit Insurance Corporation.[3] Indeed, in its concern for the safety of insured deposits,

6. Prentice-Hall, *Federal Control of Banking* (1951), p. 2445.

7. For example, the Chase National Bank - Bank of Manhattan direct merger under the latter's charter would not have been possible until 1950. Before that date, Chase would have had to go into liquidation, and the Bank of Manhattan would have had to buy Chase shares for cash.

8. Cf. Report No. 1083, 81st Congress, 1st Session, HR, to accompany HR 1161. Even before the passage of the Consolidation Act, the McFadden-Pepper Act, and the "two way street" Act, indirect consolidation was possible through a purchase merger in which the purchased bank went into voluntary liquidation.

9. Cf. Prentice-Hall, *op. cit.*, pp. 2438–50; John M. Chapman, *Concentration of Banking*, chap. III; Federal Reserve Bank of Philadelphia, *Business Review*, Nov. 1954, p. 12.

1. The right of dissenting shareholders of the absorbing institutions to demand cash payments for their shares was eliminated from national banking legislation in 1952. Cf. Comptroller of the Currency, *op. cit.*, 1952, p. 5.

2. Such mergers must be approved by the Federal Deposit Insurance Corporation, the Board of Governors of the Federal Reserve System, or the Comptroller of the Currency. This provision was introduced in 1950. Cf. Federal Deposit Insurance Corporation, *Annual Report*, 1950, pp. 155–56.

3. Cf. Prentice-Hall, *op. cit.*, p. 7709.

the Federal Deposit Insurance Corporation sometimes actively encourages bank mergers by sponsoring the merger of a weaker bank with a stronger bank.[4]

Branch banking legislation and mergers. Acquired banks were converted into branches of the acquiring banks in 85 per cent of the mergers in 1954. By these conversions the overwhelming majority of recent bank mergers came under the jurisdiction of branch banking legislation.[5] In addition, in acquiring banks and converting them into branches, state member banks must receive approval from the Federal Reserve System, and state, nonmember insured banks must secure the approval of the Federal Deposit Insurance Corporation.[6] State laws generally restrict branch banking in two ways: geographical limitations on branch banking, and minimum capital requirements for head offices and branches. Geographical limitations for national banks are the same as those for state banks in the state where the national banks are located.[7] Minimum capital requirements for national banks have not been so closely tied as geographical limitations to the standards set by each state, but recent legislation (1952) has lowered some capital requirements and moved them closer to individual state levels.[8] Although state laws often provide additional criteria to guide state officials in approving or disapproving branch applications, the National Banking Act provides no detailed criteria for the Comptroller.[9] However, the Comptroller has evolved administrative regulations for branch applications. These include the financial history and conditions of the bank, the adequacy of capital, future earnings prospects, the character of management,

4. Cf. Federal Deposit Insurance Corporation, *Annual Report*, 1949, pp. 3–5. In recent years, defalcation has often been the source of difficulty for the "weaker" bank. Twenty mergers were effected with the aid of the FDIC during the period 1945–1951. Cf. House of Representatives, 82d Congress, 2d Session, *Bank Mergers and Concentration of Banking Facilities*, A staff report to Subcommittee No. 5 of the Committee on the Judiciary, p. 57.

5. The retention of branches acquired through merger is generally also subject to the approval of banking authorities. Some exceptions to this general rule were made by the McFadden-Pepper Act, and the branch banking laws of various states also provide for some exceptions.

6. Cf. Board of Governors, *Compilation, op. cit.*, p. 3, and Public Law 543 (July, 1952).

7. Cf. Board of Governors, *Compilation, op. cit.*, p. 2.

8. Comptroller of Currency, *op. cit.*, 1952, p. 5.

9. The National Banking Act provides that the establishment of national bank branches must be "subject to the restrictions as to location imposed by the law of the State on State banks." Cf. Board of Governors, *Compilation, op. cit.*, p. 2.

the number and location of existing branches, the proposed location of new branches and distance from head office, nearest banking facilities, convenience and needs of the community served by the new branch, nature of the potential clientele, and possible business available.[1] Mostly these are technical matters. However, considerations of the number and location of branches are not solely technical, but also carry implications about the competitive effects of branches acquired in mergers. On the other hand, the "need" criterion for a branch may lead to entry by acquisition rather than by *de novo* branches, because of official apprehension about an "overbanked" community. In fact, some state laws (e.g., New Jersey, New York) prohibit *de novo* branches under certain circumstances where entry by acquisition would be permitted.[2]

Federal antitrust laws. Both the Sherman Act and the Clayton Act potentially affect bank mergers. The Sherman Act forbids every combination in restraint of trade or commerce. As of the present, however, the Sherman Act has never been invoked in a bank merger. By contrast, the Clayton Act has been invoked once, but only against a bank holding company. The Board of Governors of the Federal Reserve System is empowered to administer certain provisions of the Clayton Act for the banks of the country. Section 7 of the Clayton Act prohibits a corporation from acquiring *stock* of another corporation, if the acquisition substantially lessens competition. However, Section 7 did not apply to mergers by acquisition of assets. In 1950, an amendment to Section 7 broadened the original prohibitions against stock purchase to include acquisition of assets. Since this amendment applied only to corporations under the jurisdiction of the Federal Trade Commission, only the terms of the old Section 7 apply to banks today.[3] The Board of Governors had, indeed, urged that banks be included in an extension of the antimerger provisions of the

1. Prentice-Hall, *op. cit.*, p. 2141. In addition, by the terms of the Federal Reserve Act, state member banks are permitted to operate branches " . . . on the same terms and conditions and subject to the same limitations and restrictions as are applicable to the establishment of branches by national banks . . ." Cf. Board of Governors, *Compilation, op. cit.*, p. 3, and Public Law 543, July, 1952. The guides used by the FDIC in passing upon applications for branches by state, nonmember insured banks are the same as those for determining the eligibility of such banks for insurance. Cf. Bd. of Governors, *Compilation, op. cit.*, p. 3.

2. *Ibid.*, pp. 17, 20, 24, *passim.*

3. Cf. *Bank Mergers and Concentration of Banking Facilities, op. cit.*, p. vii. However, this provision does not apply to banks directly because banks cannot buy other bank stocks.

Clayton Act, but it was deemed " . . . impracticable to include within the scope of the act corporations other than those subject to regulation by the Federal Trade Commission."[4] The most powerful restriction on bank mergers — the prohibition against interstate branch banking — is not part of federal antitrust legislation. The failure of existing legislation to regulate bank mergers for their market effects has been noted by W. M. Martin, Jr., chairman of the Board of Governors of the Federal Reserve System:

> . . . present statutory provisions with respect to bank mergers and consolidations do not provide effective means for the prevention of mergers and consolidations which may result in a lessening of competition or a tendency to monopoly. Even in those cases in which consent is necessary, the law does not specifically require the Comptroller of the Currency, the FDIC, or the Board, in determining whether to grant or withhold consent, to consider whether the proposed merger or consolidation will lessen competition or tend to a monopoly or to consider any other specific standards or guides.[5]

Summary. Much bank merger legislation of the last few years has made it easier, not more difficult, for banks to merge. The "two way street" law, the elimination of the rights of dissenting stockholders of the absorbing bank to demand cash for their shareholdings, and the lowering of capital requirements for certain branch offices could not realistically have been expected to decrease the number of bank mergers in recent years. Although banking is highly regulated, bankers need not be so concerned as other business men about antitrust restrictions in proposed bank mergers. Neither the Sherman Act nor the Clayton Act has been effectively invoked against bank mergers. Moreover, the specific antimerger provisions of federal antitrust laws in banking do not cover mergers of banks by purchase of assets. Effective federal antimonopoly considerations enter only through the back door, primarily the geographical limitations on branch banking and various administrative regulations of the Comptroller and state officials. The legal restrictions to bank mergers are mostly technical, e.g., protection of stockholders, protection of bank capital, etc. Perhaps, too, the multiplicity of approving agencies might be regarded as an unintentional inhibiting factor in bank mergers. This gap in federal legislation led a staff report on bank

---

4. *Ibid.*
5. Letter to the Chairman of the Committee on the Judiciary, dated June 26, 1952, from the Chairman of the Board of Governors of the Federal Reserve System. *Ibid.*, p. 25.

concentration by the Committee on the Judiciary to conclude that " . . . existing laws relating directly to bank mergers . . . are inadequate to protect the public interest against the elimination of competition in banking and the centralization of financial power."[6]

IV. Effects of Mergers upon Banking Markets

Three interest groups are principally affected by bank mergers — stockholders, depositors, and borrowers. Mostly, official concern about mergers has centered on their possible effects upon borrowers, and commercial borrowers in particular.[7] Accordingly, a statute was proposed in 1952 to " . . . amend existing law so as to require all bank mergers to receive the scrutiny and approval of Federal banking authorities who in passing thereon, would be obliged to determine whether the effect of such a merger might unduly lessen competition or tend to create a monopoly in the field of banking."[8] It is the purpose of this section to suggest an analytical framework for determining the market structure effects of bank mergers.

The competitive effects of bank mergers on business loan markets[9] cannot be determined on some over-all basis, because these effects are not uniform for all bank mergers. The most important factors in analyzing the market structure effects of bank mergers can be summarized under the following headings: (1) the particular customer loan market in which the mergers take place, (2) the structure of banks participating in the mergers, (3) the absolute and relative sizes of the merging banks, (4) the comparative composition of the merging banks' portfolios, and (5) the immediate impact versus the cumulative effects of bank mergers. Each of these factors will be examined in turn.

---

6. *Ibid.*

7. For example, a staff report of the House Judiciary Committee expressed its concern about "The trend toward larger and larger banks and the elimination of many small banks which are the principal source of funds for small borrowers . . . in view of possible consequences upon the future financing of commercial enterprise." *Ibid.*, pp. 37–38. Actuated by this concern, the staff recommended " . . . the introduction of measures designed to halt the unrelenting merger trend in the field of banking, especially insofar as such mergers may adversely affect the competitive structure of the banking economy and tend to encourage undue concentration of financial power." *Ibid.*, p. 49.

8. *Ibid.*

9. Nonbank sources for business loans can, generally speaking, be ignored without serious consequence to the structure or results of the analysis. Cf. Alhadeff, *op. cit.*, chap. II.

Particular markets in which mergers occur. The market for business loans is not a single, unified market. It has been shown elsewhere that the market for business loans can be meaningfully subdivided by size of borrowers.[1] The credit standing and importance of large borrowers gives them access to a national credit market with a large number of banks. Accordingly, the large borrowers' market is reasonably competitive both in terms of credit availability and interest rates. At the opposite extreme, the credit standing of small borrowers is very limited, and they are effectively tied to their local banking markets, which are oligopolistic at best and frequently tend towards monopoly. The intermediate size business borrowers are neither small enough to be effectively tied to local banks nor sufficiently large to command easy access to the national credit market. Middle-size borrowers are an amorphous group and operate in markets which are oligopolistic at one extreme and shade from varying degrees of oligopoly to semicompetitive markets.

The competitive effects of mergers depend upon how the mergers affect concentration in particular banking markets. Because of the large number of banks in the large borrowers market, no single bank can affect market price significantly by its own action. An intensification of the merger movement would not alter the present competitive character of this market, because legal prohibitions against interstate branch banking and various geographical limitations on intrastate branch banking effectively limit the number of banks which can be eliminated by mergers.[2] The *objective* number of alternative sources confronting the small borrower is, in most instances, oligopolistic at the very best; the *effective* number of alternative sources is generally even more limited. It is evident that borrowers in the small borrower market are not affected by a merger which reduces the objective number, but not the effective number, of alternatives for particular borrowers.[3] Mergers which eliminate the effective alter-

---

1. Cf. David A. Alhadeff, "The Market Structure of Commercial Banking in the United States," this *Journal*, Feb. 1951.

2. The most important qualification to the foregoing concerns the possible effects of holding companies, or what is sometimes called group banking. Although individual banks, even national banks, cannot cross state lines to open banking offices, bank holding companies can and do acquire control of banks located in different states. Whether such holding companies could or would reduce the effective number of alternative sources for large borrowers until the market structure of the large borrower market became oligopolistic still requires much study for a definite answer.

3. Suppose, for example, that the merging banks are both large wholesale banks which do not deal to any great extent, if at all, in the small borrowers' market.

natives make the market more intensely oligopolistic, or possibly even monopolistic. In either case, the small borrower is not likely to face significantly different price results. However, if the eliminated bank had been the sole effective source of supply, the small borrower might be deprived of credit availability by the merger. This might be true even if the merged bank is succeeded by a branch of the acquiring bank, if the bank relationship established with the merged bank is not transferable (or only with great delay) to the successor institution.

Mergers in oligopolistic middle-size borrowers' markets (e.g., a large department store with operations limited to a single city) intensify the oligopolistic concentration, but probably have little effect on rate of interest and credit availability. Mergers in semi-competitive intermediate-size borrowers' markets (e.g., a manu-facturer with business outlets in a few cities, but one not large enough to enter the national market) could convert a semicompetitive situa-tion into an outright oligopolistic market, with probable results traditionally associated with this kind of switch.

Structure of merging banks. Each merger among unit banks elim-inates one objective source of credit, if the continuing bank remains a unit bank after the merger. The consequences are more complicated for branch bank mergers with unit banks or other branch banks. For example, a branch bank can discontinue outright an acquired unit bank, an acquired branch office of a branch bank, or the head office and all branches of an acquired branch bank. The discontinuance of the unit bank reduces the objective number of alternatives by one in all banking markets. The discontinuance of an acquired branch may or may not affect the objective number of alternative sources. The closing of a banking office acquired by purchase of a single branch but not its parent bank[4] would not affect the objective alterna-tives in the large borrower market. If the parent bank continued in existence, the elimination of the banking office would not affect the middle-size borrowers' market either. The closing of one banking office might not reduce the effective alternatives for small borrowers in the merger town, if the selling bank had other branches in the same (or nearby) town or city. Objective alternatives for the small

4. Separate sale of a branch to another bank is not technically a bank merger, and such sales do not appear in merger figures or in the 1953–mid-1954 data. Notwithstanding, individual branches are sold, and their effects should be recognized.

borrower would certainly be reduced by the loss of the office, if the parent bank had no other branches in the area. Finally, the number of alternatives for small and middle-size borrowers is reduced by closing all branches including the head-office of an acquired branch bank, but the effect on the large borrowers is imperceptible.

Another possibility in a merger is for a branch bank to absorb a unit bank or another branch bank and convert the acquired bank into its own branch or branches. The large borrowers' sources would not be perceptibly affected by this merger, but sources for the middle-size borrowers would be more restricted. The merger might either reduce or leave unchanged the objective number of alternatives for the small borrower. The small borrower's objective alternatives would be reduced, even though the old branch was continued by the acquiring bank on the same premises and with substantially the same personnel, if the acquiring bank already had branches in the acquired bank town. This is because a multiplicity of offices of the same bank in a town or city cannot be confused with a multiplicity of alternative sources for a small borrower. The objective number of alternatives for the small borrower is left unchanged, though the composition is different, if the converted branch is the acquiring bank's initial entry into the acquired branch's town.

It is interesting in this connection to examine the pattern of the actual mergers during 1953–mid-1954, and to note the effect upon alternative sources in the small borrowers' market. During this period 208 banks, both branch and unit, were acquired by other banks. Since the banking market for small borrowers is basically a local market, the acquisition of branches outside the head-office city of absorbed banks must be counted as separate mergers.[5] By this analytical count for small borrower markets, the actual number of 208 mergers is increased to 244. Twenty-nine banks, branches, or head-offices were discontinued outright, and the number of alternative sources for small borrowers was reduced accordingly. The remaining 215 banks, branches, or head-offices were converted into branch offices of the acquiring bank. However, as noted earlier, the conversion of an acquired bank or branch into a branch of the acquiring bank does not automatically leave unchanged the number of objective

5. In this count, two or more branches (or the head-office) of the acquired bank in a single town or city would count as only one merger, because the multiplication of branch offices of the same bank in a town does not constitute bona fide alternative sources of supply.

alternatives for the small borrower. In the period in question, 180 of 215 conversions took place in towns or cities in which the acquiring bank had no previous office. Hence, in 180 cases, or about 74 per cent of the "mergers," the number of objective alternatives facing small borrowers in local markets was unchanged. The number of objective alternatives was reduced, however, in 35 cases, for the acquiring bank already had a branch or head-office in the acquired bank (or branch) town. In 64 cases, therefore (29 discontinued plus 35 converted in the same town or city where the acquiring bank was already established), the number of objective sources of credit available to small borrowers was reduced.

Size of merging banks. The size of the merging institutions must also be considered in assessing the market structure effects of mergers. It is generally conceded that, with the exception of branch banks, size of bank is highly related to size of borrower. Roughly speaking, therefore, the different banking markets defined by size of borrower are served by banks of appropriate size in the different banking markets. In the light of the analysis presented earlier, classifications by size of bank could be listed, with probable market effects for each. For example, the acquisition by a large unit bank of a small unit bank in a different city, and the conversion of that bank into a branch of the acquiring bank, would not perceptibly improve the situation in the large borrower market, nor would it affect the number of alternative sources in the middle or small borrowers' markets.

One of the most interesting possibilities in terms of alternative sources concerns the merger of small unit banks. Consider, for example, a bank holding company with a majority stockholding or outright ownership of twenty small unit banks in a given geographical area, say a state. The merger of twenty banks, and their conversion by the holding company into a single branch system, would not perceptibly affect the large borrowers' alternative sources. If the merging banks were in different towns or cities, the local alternatives for the small borrowers would also not be affected — the local bank would simply have undergone a change of name. The small borrowers' alternatives would be reduced by the merger of banks in the same town only if the holding company had been a silent stockholder, and not a policy co-ordinator. On the other hand, the middle-size borrowers' market might actually benefit from the mergers, by an increased number of alternative sources. The market situation of middle-size borrowers is not materially affected by the objective

number of small banks, especially when the latter are widely distributed. However, the merger of numerous small banks might produce a bank large enough to serve the credit needs of middle-size borrowers. In a semicompetitive middle-size borrower market an additional source of supply might have no effect on rates, though it might stimulate increased product competition. An oligopolistic middle-size borrower market would still be oligopolistic after the mergers, and any rate or availability effects would depend on whether the new bank adopted an aggressive policy, at least initially.

The size pattern of the actual mergers during the period 1953–mid-1954 is shown in Table VI. The dominant pattern was acquisition by large and middle-size banks of considerably smaller banks.

TABLE VI

COMPARATIVE SIZE OF MERGING BANKS FOR EACH MERGER, 1953–MID-1954

Absorbed Bank Deposits/ Acquiring Bank Deposits (%)	Total Number of Mergers	PERCENT OF TOTAL MERGERS BY SIZE OF ACQUIRING BANKS[1]		
		Small Banks	Middle-size Banks	Large Banks
10.0 or less..........	104	1.9	16.5	32.0
10.01– 20.0.........	30	2.4	9.7	2.4
20.01– 30.0.........	16	3.9	3.4	0.5
30.01– 40.0.........	13	2.9	3.4	0.0
40.01– 50.0.........	8	0.5	2.9	0.5
50.01– 60.0.........	8	1.5	1.5	1.0
60.01– 70.0.........	7	1.5	1.5	0.5
70.01– 80.0.........	3	1.0	0.5	0.0
80.01– 90.0.........	5	1.0	1.0	0.5
90.01–100.0.........	2	0.5	0.5	0.0
100.01 and over......	10	2.9	1.9	0.0

Source: See note 2, p. 504. Two cases were omitted because of incomplete data.
[1] The size categories, in terms of deposits, are those used by the Federal Reserve Bank of Philadelphia, *Business Review*, Sept. 1954, p. 3. Small banks are those with $10,000,000 deposits and less; middle-size banks, $10,000,000–$100,000,000; large banks, $100,000,000 and more.

Large or middle-size banks acquired banks one-tenth or less of their own size in 48.5 per cent of all mergers. In three-fifths of the mergers, large or middle-size banks acquired banks one-fifth or less of their own size. In other words, most mergers were a kind of "mopping up" operation by large and middle-size banks. A minor pattern revealed by the table is that about 20 per cent of the mergers occurred between and among small banks.

Comparative portfolios of merging banks. Since banks are multi-product firms, with different portfolio composition, it is entirely possible for two banks in the same area to engage in essentially non-

competing types of business. For example, wholesale banks lend mostly to prime and large business borrowers; retail banks make smaller business loans and consumer and mortgage loans. The merger of a wholesale and a retail bank would have no competitive implications for the large borrowers market, nor would it significantly affect the small business borrowers formerly served by the retail bank. The only important qualification to these results would follow from a possible change of emphasis in the combined portfolio of the new bank, compared with the portfolio composition of the banks before the merger. Here, again, as the reader can observe, the effects on different loan markets would not likely be the same.

Immediate impact versus cumulative effect. In banking markets as in industrial markets, the competitive impact of individual mergers must be distinguished from the impact of cumulative mergers over time. The cumulative effect of mergers in the large borrower market is probably not more damaging than any single merger — at any rate, under present prohibitions against interstate branch banking.[6] The cumulative effect of mergers in the small borrower market would also not differ greatly from the impact of individual mergers. Since a high proportion of small borrower markets are duopolistic or semi-monopolistic, there is not much scope for the cumulative impact of mergers to deteriorate the situation. Oligopolistic small borrower markets might become duopolistic or semimonopolistic, but mergers only rarely result in bankless towns.[7] The cumulative effect of mergers would probably be felt most strongly in the middle-size borrower markets. The cumulative effect of mergers by small borrower banks might convert an oligopolistic middle-size borrower market into a semicompetitive market. By contrast, the merger of middle borrower banks or of large borrower banks could, over time, convert a semicompetitive situation into an oligopolistic one.

Market structures and market results. The analysis of this section has stressed primarily the market structure results of bank mergers. It is important to mention that a direct relationship does not exist

6. Even under present prohibitions against interstate branch banking, the operations of bank holding companies might increase effective concentration more than might be expected if attention is focused upon the formal legal independence of the banks involved. At least for purposes of this paper, however, the competitive effects of bank holding companies are moot.

7. While 321 national banks went out of operation between January 1, 1946, and December 31, 1953, only nine communities were left without banking facilities as a result. Cf. Comptroller of Currency, *op. cit.*, 1953, p. 9.

between certain market structures and particular market perform-
ance. This is conspicuously true of oligopoly markets. It is generally
presumed that oligopoly markets lead to price and output results
approximating those of monopoly. On the other hand, it is also
recognized that " . . . several influences may cause oligopoly market
behavior to depart widely from monopoly price and output policy."[8]
Hence, both market structure and market performance results are
not likely to be the same for all mergers. It follows that if public
policy is primarily concerned with the market effects of bank mergers,
the legislation and administration invoked to put that policy into
effect must be highly discriminating. All bank mergers cannot be
painted with the same brush.

V. Conclusions

1. Bank mergers have increased at a sharply accelerated rate
during 1953 and 1954, compared with the merger rates of the forties.
2. Two basic factors seem to have motivated the sharp accelera-
tion of the bank merger rate:
   (a) pressures resulting from uneven rates of growth by different
       banks during the rapid over-all expansion of the last few
       years.
   (b) the desire to establish or expand branch bank systems.
Although management problems may have been involved in the
recent upsurge of bank mergers, they were not a major initiating
factor. Similarly, higher operating costs and lower earnings did not
exert a systematic pressure upon banks to merge. The limitations
on bank lending imposed by the 10 per cent rule are also not a per-
suasive explanation for the sudden acceleration in the over-all rate
of mergers in 1953 and 1954. Finally, the fact that banks are often
worth more dead than alive is a facilitating condition, not a primary
cause, of recent bank mergers.
3. Since the Consolidation Act of 1918, much federal legislation
has made it easier for banks to merge within and across jurisdictional
(chartering) boundaries. Under present laws, bank mergers are also
not seriously impeded by antitrust considerations.
4. The major factors which influence the market results of bank
mergers are the particular markets in which mergers occur, the struc-

8. J. Fred Weston, *The Role of Mergers in the Growth of Large Firms* (Berkeley
and Los Angeles: University of California Press, 1953), p. 108.

ture of merging banks, the size of merging banks, the comparative portfolios of merging banks, and the immediate impact versus the cumulative effect of mergers.

5. The market structure and market performance results of bank mergers vary widely. Accordingly, legislation and administration concerned with the market effects of bank mergers must be highly discriminating.

CHARLOTTE P. ALHADEFF.

BERKELEY, CALIFORNIA

DAVID A. ALHADEFF.

UNIVERSITY OF CALIFORNIA
BERKELEY

PRODUCT DIFFERENTIATION AND
WELFARE ECONOMICS*

By ALEX HUNTER

I

There is general agreement that the study of welfare economics should result in some proposals for practical policy. But so far, in the important field of imperfect and monopolistic competition, very little has emerged since the one or two pioneer articles which, at the time, appeared to indicate great possibilities. It is the purpose of this article to show why this is so; to show that the form of modern welfare economics is incompatible with practical applications of policy in this field; to demonstrate that welfare economics is a methodological system rather than applied theory and that when used as theory it can be seriously misleading on matters of economic welfare. To do this the article analyzes the difficulties which arise out of the welfare aspects of product differentiation.

Professor Chamberlin's work has made product differentiation an old story. But certain points which bear on our topic require repeating here. Product differentiation can be interpreted in a strict sense to mean that, economically (and this includes an important welfare sense as will be seen presently), even physically similar goods are not the same to the consumer if there exist quite small variations of a qualitative nature. Such differences need not include the exact physical make-up of the good itself; although such factors as technical merit, finish and quality of material are, of course, important. The differences need only exist in the circumstances surrounding the sale of the good. Thus trade-names, labels, packing and retail services are some of the factors which must, for most cases, be con-

sidered to be a minimum part of the consumer's assessment of a good. The extreme case is where the homogeneity of a good, identical in all other respects, is broken down so far as the consumer is concerned if there is even a minimum differentiation of location and he has to pay something either in cash, or inconvenience, because of it. To put the proposition succinctly, product differentiation becomes an economic attribute of *any* good in its retail stages because product differentiation is in the very nature of retail organization. Full homogeneity of product can only be assumed to be present when every one of these factors (location, packing, service, etc.) is the same.[1]

A possible argument which can be made against this interpretation of the economic world is that the degree of differentiation is so slight in most cases that it makes little difference to the perfectly competitive hypotheses.[2] Or in schematic terminology, elasticities of demand are so high as to be indistinguishable from the infinite elasticities posited by perfectly competitive theory. This may be so, and perhaps no final answer can be given without actual measurement. But it is difficult to think of an example in which anything approaching the perfect conditions holds. It is difficult, for instance, to conceive of a retailer who would lose all or even most of his sales if he raised his price a little; or who would secure all or even most of the market if he reduced his price a little.

A more important difficulty for this approach perhaps lies in the existence of raw material markets with commodity exchanges which do, more or less, conform to the conditions of perfect competition analysis. But their existence is, generally speaking, an irrelevancy for the question of consumers' welfare. For it is only in the raw material stage that these perfect conditions are found at all. The end-product reaches the consumer through some kind of retail establishment, professional traveler, etc., in which case the end product is always subject to the monopoly influences of location, service, packing, and other forms of product differentiation. That is, the *consumer* is always subject to monopolistic influences.

(Besides the above arguments for the ubiquity of product differentiation which are derived from purely static considerations, the point could also be pressed that product differentiation is an indispensable agent in the dynamics of competitive effort, as we understand it in a modern society, where the entrepreneur is everywhere

1. Cf. E. H. Chamberlin, *Theory of Monopolistic Competition*, pp. 62–63.
2. Cf. J. H. Hicks, *Value and Capital*, pp. 83–84.

expected to pursue the introduction of new materials, methods and commodities. To quote J. A. Schumpeter's phrases on this point: "The bulk of what we call economic progress is incompatible with [perfect competition]"; and "in this respect perfect competition is not only impossible but inferior, and has no title to being set up as a model of ideal efficiency."[3] However, this analysis will be confined to arguments within a static framework. The intention is to criticize welfare economists on their own ground rather than plead irrelevancy of this particular nature, however sound the grounds.)

II

The picture which emerges from the analysis of the effects of product differentiation appears to indicate a world in which competition will be at the very least, of a monopolistic variety. Yet one would suppose, prima facie, that welfare economics is not competent to judge situations of this nature since, by reason of its form, it is capable of making only special kinds of statements — statements which in fact rest upon the highly abstract assumptions of perfectly competitive conditions.

Although welfare economics is not simply perfect competition, the end result certainly does have a strong family resemblance to the perfect market situation.[4] This can be seen in the main set of conditions of the welfare optimum — what Professor Hicks calls the marginal conditions — which state that the marginal rates of substitution (for exchange) and the marginal rates of transformation (for production) between any two commodities must be the same for every individual or firm which consumes or produces them both. This is in order to secure, on the one hand, as complete satisfaction as possible in the circumstances for the consumer, and on the other hand, to secure an allocation of resources which is correct for the particular pattern of wants. The perfect market here provides a ready-made model for the welfare economist's purposes. In a perfect market the price of the good is the same for every individual. And every individual is enabled to maximize his position by extending his buying or selling activities to the point where the ratio of every pair of prices becomes the same as the relevant rates of substitution and trans-

3. *Capitalism, Socialism and Democracy*, pp. 105–6.
4. The strong family resemblance no doubt arises from the close historical association between analytical perfect competition and welfare economics. Cf. H. Myint, *Theories of Welfare Economics*, Part II and especially p. 91.

formation. The price-equal-to-marginal-cost rule is a derivative condition of the marginal conditions and it is noteworthy that it also is common to both welfare economics and perfect competition.

The analogy serves to draw attention to this fact that the statements and recommendations of welfare economics on practical matters depend on business men *behaving* as though actually confronted by a perfect market and reacting in the welfare manner to make price equal to marginal cost.[5] From the analysis of section I it appears most unlikely that many markets are of such a type. If so, the price-equal-to-marginal-cost rule reduces considerably the usefulness of welfare economics for practical policy.[6]

To be more precise on this important issue, product differentiation has the effect, for all end products, of confronting the business man with a demand curve for his good which is tilted from the horizontal. To maximize his gains he is compelled to charge a price which, in some degree, is greater than marginal cost. Hence the productive units of the economy *cannot* conform to the behavior demanded of them by the welfare system: they cannot make price equal marginal cost. And there is with unrestricted product differentiation a distribution of resources which, if every factor moves to the firm where its reward is highest, certainly will not (and should not) coincide with the distribution desired by consumers *as given by the perfectly competitive criterion.*

This is a very serious difficulty for a field of study which is ambitious for practical applications of its work. It might be overcome if what can be called the commodity conditions of perfect competition were implemented as part of the welfare policy (those conditions requiring perfect homogeneity of goods in respect of appearance, quality, service, etc.). This would imply the need for a standardized end product in all goods. But common sense revolts against such a

5. The marginal conditions are, of course, far from consisting of a complete welfare system. There are many other equally important conditions in the abstract model — that the distribution of income be at an agreed optimum, that the marginal rate of substitution of one kind of work for another kind be equal for all individuals who perform both types of work, etc. For the complete set of welfare conditions see A. P. Lerner, *Economics of Control*, I. M. D. Little, *A Critique of Welfare Economics*, or M. W. Reder, *Studies in The Theory of Welfare Economics*.

6. It has on occasion been urged that price proportional to marginal cost is an adequate rule here — something less rigorous than the perfect competition model. But it has been demonstrated, e.g., by Lerner and Reder, that the only price proportional to marginal cost which can hold for all products simultaneously is that of unity (i.e., the case of price *equal* to marginal cost). Cf. Lerner, *op. cit.*, pp. 102–5, Reder, *op. cit.*, p. 42, and P. A. Samuelson, *Foundations of Economic Analysis*, pp. 239–40.

solution. Product differentiation is after all a method of satisfying more accurately and more fully the varied demands of the consumer. There is a presumption in the democratic society that the consumer is entitled to this diversification if he so desires.

This brings us to the crux of our critique. Product differentiation is undoubtedly part of what the consumer considers to be his economic welfare. Yet, because it prevents the *practice* of marginal cost pricing, it prevents the implementing of the purely "efficiency" conditions of welfare economics. No recommendation or policy — apart from the not very practical possibility of subsidies all round — can compel the business man to conform to a price-equal-to-marginal-cost rule. One is led to question the claim that welfare economics can be an adequate "means of criticizing or testing the efficiency of production by private enterprise."[7] The dilemma for theoretical welfare economics appears to be whether it should leave product differentiation out of its analysis, and consequently ignore an important aspect and application of welfare; or include it and make welfare economics cumbrous and difficult to handle. One begins to suspect that the structure of modern welfare economics is selected[8] to suit the techniques of theory rather than the practical requirements of welfare criticism.[9]

It is not suggested that welfare economists are unaware of the difficulties created by the monopoly elements contained in product differentiation. That the problem is tackled at all is sufficient evidence to the contrary. But it does seem as though the attractions of a determinate and tidy solution are so great that it leads to the adoption of a version of economic welfare incompatible with what the consumer understands by that term. It will be suggested in the course of this article that the correct procedure is to recognize that an essential element of consumer welfare is contained in the almost universal preference for product differentiation of goods — and that welfare analysis should begin at this point.[1] To illustrate the argu-

7. J. R. Hicks, "Foundations of Welfare Economics," *Economic Journal* 1939, p. 706.
8. It is interesting to note that Professor Boulding in making substantially the same criticism of welfare economics attributes this weakness to the limitations of indifference curve analysis. (*Survey of Contemporary Economics*, II, 12.)
9. Cf. E. H. Chamberlin, "Product Heterogeneity and Public Policy," *American Economics Review*, May 1950, p. 89, on this dilemma.
1. It should be noted that Chamberlin has always been quite clear on this point, that the basis of welfare economics is faulty so long as it disregards the problems posed by variety. Cf. *Monopolistic Competition*, 6th ed., p. 214 (or this *Journal*, Aug. 1937, pp. 576–77); also, "Product Heterogeneity and Public Policy," *op. cit.*

ment two important articles will be discussed which relate welfare economics and imperfect competition.

III

There is first of all Professor Hicks's work along these lines.[2] It should be noted that the main structure of his "Foundations" article is composed of "marginal" conditions in which "the marginal rate of substitution between any two commodities must be the same for every individual (who consumes them both) and for every producing unit (which produces them both)." That is to say, this Paretian type of optimum favored by Hicks and the "new" welfare economics, in which no one can move to a position he prefers without at the same time moving someone else to a position which is less preferred, is defined in its major characteristics by the conditions which, as we have already indicated, are the equivalent of analytical perfect competition. Hicks's other sets of conditions are brought in as refinements of this essential structure. The "stability" conditions ensure that the marginal conditions are established at positions of maximum and not minimum satisfaction. And the "total" conditions are concerned with the possibility that greater satisfaction may be created by the abandonment of an existing product or the introduction of a new one.

The claims made for these "foundations" are sweeping. "No economy has ever existed nor (we may be sure) will ever exist, to which they are irrelevant."[3] It is therefore all the more interesting to consider the treatment given to product differentiation within this system.

The form of treatment emerges in a consideration of the problem of the optimum number of firms in a monopolistically competitive industry. Linking it up with the total conditions of his optimum system Hicks asks the question whether a reduction in the number of products would be conducive to a movement towards the optimum.

The analysis here turns on a feature of imperfect and of monopolistic competition — that the marginal resources and therefore all others are not receiving the full value of their marginal product: this in his terminology varying with "the degree of monopolistic exploitation."[4] There is as a result a producers' surplus even at the

2. "Foundations of Welfare Economics," *op. cit*, p. 696.
3. *Ibid.*, p. 706.
4. The "exploitation" thesis is held generally by those economists who follow the Joan Robinson school of Imperfect Competition; it is, however,

margin of each of the imperfectly competitive firms which may out-weigh, and therefore compensate for, the loss to consumers (the consumers' surplus) and to the producer (his excess of earnings in this use) if certain firms of the industry are closed down. The condition for the continued existence of a firm is "that the sum of consumers' and producers' surpluses generated by its activities must be greater than the producers' surplus which would be generated by employing its factors (and exploiting them) elsewhere."[5] Hicks puts forward this use of Marshall's surplus declaring that it provides a "general rule" for the application of welfare principles to imperfect competition. The "rule usually given" (that of price-equal-to-marginal-cost) is a special case of this rule and holds only where entry is free, and where the products of the firms are "very close" substitutes, or are distinguished only by irrational preferences.

The first comment appropriate to this surplus technique for imperfect competition is that it appears to introduce an internal contradiction into Hicks's welfare system. It seems, using this surplus analysis, that some firms will be permitted to continue to produce at positions of price greater than marginal cost. This could (frequently) occur for instance where there are genuinely "rational" preferences for a differentiated product creating a consumers' surplus sufficient to justify the retention of a particular variety; or where the varieties of a monopolistically competitive industry, while similar to one another, are yet not "very close" substitutes for one another. But such a procedure overlooks the fact that there is a clash with Hicks's set of "marginal" conditions in which the marginal rate of substitution between any two commodities must be the same for every individual (who consumes both) and every producing unit (which produces both). These marginal conditions, as has been pointed out, involve for *every* firm in the economy, compliance with the price-equals-marginal-cost rule; otherwise there would be a faulty allocation of resources between firms, a faulty substitution of work for leisure, etc[6].

specifically denied by Chamberlin who points out that *all* factors (*including entrepreneurship*) receive less than the value of their marginal products; and that therefore all are equally "exploited." Cf. Joan Robinson, *Economics of Imperfect Competition*, p. 283; R. F. Kahn, "Notes on Ideal Output," *Economic Journal*, March 1935, p. 23; J. E. Meade, *Economic Analysis and Policy*, Part II, chaps. 4, 5; and E. H. Chamberlin, *Monopolistic Competition* (5th or 6th ed.), pp. 182, 216–18.

5. *Op. cit.*, p. 710.

6. There is the further criticism that "consumers' "surplus . . . is a toy, in the sense that it cannot provide us with any *practical objective criterion*." (Little, *op. cit.*, pp. 174–75, my italics.)

The existence of this internal contradiction in Hicks's welfare system brings out the perhaps more fundamental point that the use of surplus analysis in order to judge whether or not some products of an imperfectly competitive industry are redundant, really sets up a different criterion of economic welfare from that provided by the "marginal" conditions. The "marginal" conditions here imply a model in which individual choice is closely confined to long lines of homogeneous goods and in which (given sufficient numbers of buyers and sellers) the market price of a good is a parameter for each consumer. The surplus technique, on the other hand, implies a model in which choice contains personal elements whereby the consumer is influenced through considerations of quality, appearance, service, convenience, etc., towards the products of this or that individual producer; and these personal elements, besides breaking up the atomistic quality of the market, are judged as being creative of satisfactions which provide a special category of consumers' surplus. To repeat a point already made, economic welfare means something different in each case. It is not surprising that the conditions of one are not reconcilable with the conditions of the other.

This internal dichotomy apart, the question remains which of the two models is more fundamental to Hicks's welfare system.

There is no doubt that Hicks is sympathetic to the point of view that consumer preferences for a differentiated product *should* be considered. This is borne out by the line of argument in another context where he is examining partial welfare economics. Here most definitely he indicates the view that product differentiation can generate a consumers' surplus "by the mere fact" of making a commodity different. He further states that: "To emphasize the differences between the products for the purpose of getting a downward sloping demand curve for the individual firm, and then to neglect the differences between the products (irrational references!) for the purpose of neglecting the consumers' surplus, is both inconsistent and practically dangerous. To neglect the producers' surplus for the purpose of equating price with average cost imports quite unnecessary difficulties into the theory. These two simplifications together have been responsible for the quite incorrect idea that the number of firms in an industry ought always to be reduced when an industry is imperfectly competitive — an idea which we should at once recognize as absurd if we were not bemused by diagrams, and whose practical realisation could only lead to universal monopoly."[7] This quotation

7. *Review of Economic Studies*, VIII (1941), 116.

is more than sympathetic — it illustrates admirably a main point of this article.

Nevertheless, as already pointed out, the main structure of Hicks's welfare economics is contained in his "marginal" conditions; they are the starting point of his thought on welfare as can be seen from the significantly central position which they occupy in his statement of welfare conditions. The "stability" and "total" conditions (and the surplus analysis) are attached merely as important qualifications in order to cover difficulties which emerge from this approach. Hence the analysis of imperfect competition in Hicks's welfare economics is no more than a rather awkward appendage to a system which is based essentially on the perfect competition tradition. Significantly enough, Hicks himself states that: "The critique of private enterprise naturally begins by pointing out the one conceivable case in which an optimum position may be obtained by perfect laisser-faire. This occurs when competition is perfect in all industries. . . ."[8]

To sum up this section: Hicks's approach to economic welfare by means of the usual perfect market type of analysis encounters considerable difficulty, and actual contradiction when it analyzes monopolistic competition of the product differentiation type. The inappropriateness of the approach to a practical problem of welfare becomes explicable when it is realized that two quite different versions of economic welfare are under discussion.

IV

Our second discussion centers on Professor R. F. Kahn's important article "Some Notes on Ideal Output,"[9] which has influenced much of the thinking in welfare economics as to policies for monopoly. In this article the "inevitability" of imperfection or monopoly elements in the economy is specifically catered for; and this at first sight appears to conform to the views of this article as to the unique characteristics of product differentiation. But it is still maintained that the optimum as defined by analytical perfect competition is, nevertheless, a determinate concept by which, somehow or another, using various devices, a guide for the purpose of securing what could be called a second-best optimum is created. This in Kahn's case consists of an "average degree of imperfection," and can be summed up in the working rule of making price in the same proportion to marginal

8. "Foundations of Welfare Economics," p. 706.
9. *Economic Journal*, 1935, p. 1.

cost for all units in the economy, instead of price being equal to marginal cost all around. Hence the perfectly competitive equilibrium is a starting point for the consideration of the optimum of an imperfectly or a monopolistically competitive economy. And as Kahn sees it, although imperfection in the economy is inevitable, an ideal output can yet be determined if, as he puts it, the ratios of "marginal revenue to the individual firm and price" are equal "in the particular and alternative use" for a factor of production. (Or in the more usual and simpler version preferred in this text — that price be in the same proportion to marginal cost for all productive units.) It then follows that precisely the same condition for maximization of the national dividend as in perfectly competitive conditions still holds; that the "amount of a factor in any use will be ideal when the value of the marginal product of each marginal unit (unit situated in the margin of transfer) is the same in that use as in the alternative occupation."[1]

The device by which this condition for an optimum allocation of resources is to be created is an arrangement of all industries in the economy "in a descending order of degree of imperfection of competition." Then some form of average of imperfection is found, and by means of taxes and subsidies on the factors of production, the industries above this average are made to expand their outputs, while those below are made to contract their outputs, until such times as price is in the same proportion to marginal costs all round; or, until the average degree of imperfection exists in all industries.

No one could object to the logic of Kahn's argument. As before we are concerned with assumptions. The question which should be asked here is whether the equating of the value of marginal products between every alternative use in this particular sense really provides an "ideal output" — ideal in the sense that all aspects of economic welfare are included in its conditions.

It has already been pointed out that later and more rigorous interpretations (e.g., Lerner and Reder) of the conditions of maximum welfare have shown that a price-proportional-to-marginal-cost rule for the theory of the firm is inconsistent with a more general statement of welfare conditions. But we can leave this objection to one side for the moment in order to discuss Kahn's suggestions in respect of his attitude to product differentiation. And here we find a defect. In assuming that an ideal output requires a standard degree of deviation from what we have called the commodity conditions of perfect competition in order to secure a correct allocation of resources,

1. *Ibid.*, pp. 19–22.

Kahn is assuming away a most important element of economic welfare. Just as the consumer considers differentiation of product to be part of his welfare so also he will (presumably) wish to exercise his preference for this differentiation unrestricted by comprehensive regulations as to what extent of variety he should be permitted. The important point here is that Kahn's plan misconceives the reality of product differentiation which arises from a number of circumstances, the degree of importance and complexity of which are unique to the various individual situations. It cannot be maintained as a matter of theory *how much* product differentiation should exist in a certain field; or in the whole economy. The "a priori" element in economic analysis has no place here. As Chamberlin succinctly puts the matter, "the differentiation of product is not, so to speak, uniformly spaced; it is not distributed homogeneously among all the products which are grouped together. Each has its own individuality and the size of the market depends on the strength of the preference for it over other varieties."[2] It is clear that a uniform pattern imposed simply for the sake of an "efficient" allocation of resources does considerable violence to any common-sense concept of the welfare attributes of consumer preference.

It helps our analysis if we consider the reasoning which appears to underlie Kahn's policy. It seems to be implied that the major or the only causes of imperfection in the economy are those created by what are called "lumpiness" or "indivisibilities" in the factors of production. That is, the type of situation in which there is a factor, usually fixed capital, the units of which are so large that over a considerable range of output there may be decreasing costs. It becomes difficult in practice to make the technological output of such a productive unit — for instance one more railway line — coincide with the demand for its goods or services. This interpretation of Kahn is borne out by the structure of his article. It moves from a discussion of external economies which can be created by subsidizing "lumpiness" or "indivisibilities" in the factors of production[3] into a more general analysis of imperfection of competition in which all elements of imperfection are lumped together in the one analysis and are subjected to the "average degree of imperfection" policy outlined above. It is worth quoting the passage in which this transition takes place.[4] "It is by now fairly evident that instead of thinking in terms

2. *Monopolistic Competition*, p. 82.
3. *Op. cit.*, pp. 11–12.
4. *Ibid.*, p. 16.

of external economies and diseconomies an alternative route is open to us. In all these cases the failure of laissez-faire to achieve the best of all possible worlds may be fairly laid at the door of the imperfection of the pricing system. In an ideal world price would always be equated to marginal cost to society and all would be well.''[5] That is, imperfection of competition is to be regarded as so closely analogous to the problem of providing external economies that it may be subjected to similar taxation and subsidizing processes — this, by making ''the value of the marginal product of each marginal unit'' the same in one use as in another, raises the national dividend.

Kahn has other techniques for obtaining an average degree of imperfection, but this particular device for redistributing the hired factors of production is specifically mentioned[6] as standing on its own merits, without assistance, as a method of raising the national dividend. It can be seen that there is no suggestion in this policy that imperfection of competition, the divergence of price and marginal cost, might have origins other than that of ''indivisibilities''; and that it might require treatment different in kind from that accorded to external economies. Product differentiation is a phenomenon which has no place in imperfect competition if we are to judge from a technique which, if pushed to its logical conclusion, would succeed in compelling firms to contract and expand even though their divergences of price from marginal cost are due to phenomena quite unconnected with economies of scale.

This point becomes clearer if we examine Kahn's interpretation of product differentiation as it emerges in another of his devices for implementing the ''average degree of imperfection'' policy. He suggests that where there is imperfection of competition due to product differentiation this is again a case for taxing and subsidizing — on this occasion the entrepreneur. Imperfection of competition due to product differentiation is a situation which can be brought into line by altering the number of (rationally preferred) products in an industry. There should be fewer products or varieties where imperfection is less than the average. The entrepreneur himself is taxed or subsidized in order to compel his movement into and out of the industries concerned.[7] This technique is best seen as the complement of the other already mentioned. ''The treatment which should be applied to entrepreneurs is precisely the opposite to that which is appropriate

5. The influence of the Robinsonian school of Imperfect Competition on the above approach is discussed later, p. 547.
6. *Op. cit.*, pp. 27–28.
7. *Ibid.*, pp. 28–29.

to the hired factors. Thus taking retail selling as an example of an industry in which competition is exceptionally imperfect, we may conclude that, just as it is desirable to increase the amount of hired factors engaged in retail selling, so it is desirable to diminish the number of shops."[8]

This movement of entrepreneurs into and out of industries provides a specific instance in which both the role and nature of product differentiation is evidently misunderstood or neglected.[9] A tax or subsidy on the entrepreneur to make firms larger or smaller will alter to some extent the "imperfection" of competition in the required direction by altering the number of products within an "industry." But it is not clear that anything like an approximation to the desired "average degree of imperfection" will be obtained. The majority of product differences need not vary with the size and number of firms. For instance, product differences based on quality, technical specification, packaging, trade-marks and other individual factors could, and may be expected to, remain relatively independent of changes in the size of the firm. If they vary at all, there is no guarantee tha tthe "degree of imperfection" will move in the required direction. The opposite could conceivably happen. A retail establishment might become more imperfect (greater divergence of price and marginal cost) the larger it became if, for instance, it was enabled to spend more on advertising appeal, special services, etc. Indeed the change to larger size and fewer units in retail trade would, by itself, appear to lead to greater inconvenience of location and therefore, in this one respect, greater imperfection. Kahn's technique for the movement of entrepreneurs in order to alter the degree of imperfection does no less than attempt to alter the marginal conditions of certain aspects of phenomena which in fact have no marginal characteristics. Or, to put this in another way, his approach to the concept of imperfect competition is narrowly quantitative. The following quotation illustrates his attitude. "All this (discussion of the ideal size of the firm) has a simple interpretation in terms of common sense. Imperfection of competition *implies the existence of internal economies.* Now internal economies no less than the external economies considered in the earlier portion of this article provide a basis for interference. If the internal economies are everywhere equal no diversion is called for."[1] That is, Kahn appears to regard all "imperfection," including product

---

8. *Ibid.*, p. 28.
9. Cf. Chamberlin, *Monopolistic Competition*, 5th or 6th ed., pp. 216–18.
1. *Op. cit.*, p. 31 (my italics).

differentiation, as reducible to some form of internal or external economy — phenomena amenable to quantitative treatment — and at no point do his practical policies appear to suggest that imperfection of competition may have important *qualitative* aspects which can in no way be controlled by taxes and subsidies.

<p style="text-align:center">V</p>

It will now be making itself clear to the reader that this article is concerned with emphasizing the distinction between "imperfection" of competition directly created by the demand conditions of a commodity (product differentiation) and imperfection of competition due to cost conditions (internal economies, including indivisibilities). Both of these types of imperfection break down the particular "ideal" market conditions of competition commonly required by welfare economics — the one by creating certain relationships between buyer and seller, the other by contributing to the large size of the production unit relative to its market. There is an analytical similarity, of course. Product differentiation creates discontinuities in the market demand for a general class of goods: there is a considerable gap in type, service, quality, etc., between each variety of the commodity in this broad sense. This is much the same as saying that the groups of individuals who consume each of these varieties have, as under the "indivisibility" type of imperfection, ties of a monopolistic nature for a particular producer. That is, each type of "imperfection" in its own way gives some degree of control into the hands of the producer; each is responsible, directly or indirectly, for a demand curve of finite elasticity; and each contributes to lowering cross-elasticities of demand. Analytically, and in practice, it is difficult to decide where one begins and the other ends. But for the purpose of welfare economics it is essential to emphasize the distinction between these two sources of imperfection. They fall into quite distinct categories from this point of view. To reorganize industry in order that marginal productivities are equalized in every alternative use is a useful device; and other things being equal it is a procedure which should be pursued with respect to *given* demands as far as possible. It is an entirely different proposition to attempt to reorganize consumers' tastes into a pattern which ensures that marginal productivities are equalized in every alternative use. This is a confusion of ends and means indeed. The desire for considerable differentiation of product is as fundamental as anything can be in economic welfare; and one cannot

avoid the conclusion that if this should mean that marginal productivities cannot be equalized in every alternative use then it is a "loss" we must accept. Providing the consumer is willing to pay the extra cost there is no reason for depriving him of this privilege of having a differentiated product.

An interesting problem is why this confusion of the two types of imperfection should have occurred in the literature of welfare economics. Very probably it has some origin in the formulation of Mrs. Joan Robinson's *Economics of Imperfect Competition*.

Several factors point to this conclusion. First, it is significant that Mrs. Robinson's analysis is in terms of "industries" producing "a single commodity,"[2] an unfortunate definition of an industry considering that Mrs. Robinson's book is in fact largely concerned with the problems of heterogeneous products.[3] (Compare this with Chamberlin's "groups" producing a range of associated products.) This definition of an industry as producing homogeneous products, it should be noted, is also implicit in Kahn's treatment of an industry as a group of firms which can become more or less perfect by varying the size of the firms (more or less of the factors of production) and the number of firms (decreasing or increasing the number of entrepreneurs). Neither Mrs. Robinson nor Kahn consider product differentiation to be an obstacle to quantitative adjustment of resources and firms. Secondly, it is not surprising to find that the term "monopoly," in its usual sense of meaning control over supply, is in Mrs. Robinson's book confined to the *industry* (or to the firm only where it coincides with the industry).[4] Given her definition of an industry Mrs. Robinson is compelled to maintain the traditional dichotomy between monopoly and competition although in fact her "imperfect competition" is obviously concerned to analyze the monopoly elements of differentiated products.[5] (Compare this with Chamberlin's specific references to the monopoly influences possessed by any seller of a differentiated product).[6] Thirdly, we find further

2. *Op. cit.*, p. 17.
3. *Loc. cit.*
4. *Economics of Imperfect Competition*, Book 4 on "The Comparison of Monopoly and Competitive Output" and Book 10 on "A World of Monopolies." Mrs. Robinson in her Introduction also uses "monopolist" to mean a "single seller" (pp. 5–6). But this is evidently not intended to be an operational definition in her book; and indeed since "single sellers" are found in any market structure it could not be so used.
5. Cf. *ibid.*, pp. 5–6, 50–51, and 100–1.
6. *Monopolistic Competition* (5th or 6th ed.), pp. 7–9 and 208–9 for a discussion of these points.

confirmation of the confusion of the two types of imperfection when Mrs. Robinson appears to suggest that the "imperfection" stemming from product differentiation is merely a matter of the numbers in the market; as the number of firms (and products) in a market increases, "if the new firms were set up, so to speak, in between the old firms (either geographically or in respect of special qualities which appeal in various degrees to different customers) . . . successive increases in demand of this type would ultimately remove market imperfection altogether."[7] This fallacy we have already encountered in Kahn's analysis.[8] Altogether, Mrs. Robinson's formulation of "imperfect competition" has the result that she theorizes about what is admittedly a world of heterogeneous products as though they were homogeneous — and this with strong welfare implications for the reorganization of industry and distribution.[9]

The initial confusion gains powerful support from the "indivisibility" proposition already mentioned: it is put forward in support of Mrs. Robinson and in an aura of theoretical respectability that imperfection of competition arises only because factors are not perfectly divisible. Briefly, the argument runs that if all factors were perfectly divisible there would be no economies of scale and therefore there would be constant costs. With horizontal cost curves but downward sloping demand curves, actual and potential entry of new producers would compel existing producers to reduce output to make marginal revenue equal marginal costs and thus recover some of their lost profits. This process would continue until demand elasticities would become infinite and producers so numerous as to lose all control of their markets. Consequently, indivisibilities, or economies of scale, which give the downward slope to the firms' costs curves, are the only creators of imperfection of a true "economic" type. All other monopoly elements are the result of inertia, restrictive agreements, etc., and are "institutional" monopolies. In partic-

7. *Ibid.*, pp. 101 and 313. The relationships between product differentiation and numbers in the market is given a very full discussion by Chamberlin, *op. cit.*, pp. 196–98.

8. Cf. above, p. 545.

9. Cf. the points made above; also her chapters on "Exploitation" and "A World of Monopolies" (chaps. 25–27). One possible interpretation of Mrs. Robinson's welfare treatment is that for her *all* product differentiation is classifiable as "irrational" preference. This is nowhere made explicit in her book; on the other hand, it is an explanation which fits most of the points made above. The distinction between "rational" and "irrational" preferences was first made explicit in Kahn's article (*op. cit.*, pp. 24–26).

ular, "product differentiation will by itself never prevent the establishment of perfect competition."[1]

The fallacy of this proposition is attacked at its source in a detailed and well-argued article by Chamberlin.[2] What he shows is that the term "divisibility" usually involves a tautology: that to posit "perfect divisibility" of technique or equipment is to posit the absence of economies of scale because a divisible service is *defined* as one equally efficient at all outputs. Thus the assumption of perfect divisibility is merely another way of assuming there are no economies of scale. But the absence of economies of scale is only an abstract possibility, a theoretical excursion. It is demonstrated that efficiency[3] is functionally connected to the size of output, partly through the possibility of better organization, an increase in specialization, dexterity, etc., and partly through the possibility of using more efficient capital instruments. Hence there must be a decreasing cost section of the producer's cost curve quite apart from the question of divisibilities. Thus in Mr. Kaldor's model the influx of new producers must inevitably lead to rising costs at some point as the advantages of larger outputs are lost; and perfect competition, through the medium of constant costs, cannot occur as he suggests. Economies, in some sense or other, are always possible with increasing output up to some point, and therefore there is no reason to suppose that they have any intrinsic relationship to the presence or absence of monopolistic competition.[4]

On the other hand, the slope of the demand curve does exhibit a direct relationship to the presence of monopoly elements. Where there are large numbers involved, if the demand curve is horizontal (homogeneous products), the equilibrium is one of price equal to marginal cost and we have pure (or perfect) competition. But if the demand curve is tilted (product differentiation) the equilibrium is one of price greater than marginal cost. Since the decreasing costs

1. The main source is N. Kaldor, "Market Imperfection and Excess Capacity," *Economica*, 1935, pp. 41–43 and "The Equilibrium of the Firm," *Economic Journal*, March 1934; also, "Chamberlin on Monopolistic and Imperfect Competition," this *Journal*, May 1938, for his critique of Chamberlin's contrary point of view.

2. "Proportionality, Divisibility and Economies of Scale," this *Journal*, Feb. 1948, or Appendix B of *Monopolistic Competition*, 6th ed.

3. Pp. 239–44 of Appendix B.

4. *Ibid.*, p. 238. Associated problems treated by Chamberlin's article are the analysis of diseconomies of scale and the relationship between proportionality of factors and size of output.

section of the cost curve exists under both pure and monopolistic competition, it appears that the demand conditions, i.e., product differentiation, provide the only sufficient condition for the latter.[5]

It becomes clear that the source of confusion between "imperfect" competition due to economies of scale and "imperfect" competition due to product differentiation is indeed to be found in the formulation of *The Economics of Imperfect Competition*. The extent to which the confusion has crept into welfare analysis is directly evidenced by Kahn's important article on ideal output, already cited, and by the structure of J. E. Meade's well-known text, *Economic Analysis and Policy*, clearly and specifically based on the approach of Mrs. Robinson and Kahn.[6] Further, we find A. P. Lerner repeating the proposition that "in the absence of indivisibility perfect competition would be possible everywhere"; he is also prepared to undertake wholesale manipulation of factors in order to establish his universally applicable "Rule" of price-equal-to-marginal-cost; and where product differentiation is specifically mentioned its "problem is solved simply by keeping to the same old Rule."[7] There is a danger that the mistakes of this formulation may become perpetuated in the applications of welfare economics such are the attractions, for many analytical economists, of manipulating a universe of factors and firms according to their "marginal" values.[8] It must be concluded that

5. Cf. Chamberlin, *Monopolistic Competition*, pp. 193–94; also, R. Triffin, *Monopolistic Competition and General Equilibrium Theory*, pp. 153–55, for an alternative exposition of the same point.

6. Part 2 on "Competition and Monopoly" (chap. 6 especially) and the Preface.

7. *Economics of Control*, pp. 183–84. See also 64, 75.

8. Most recent writers appear to have followed this line of approach for applications of welfare economics. E.g., Reder would intervene in all cases of imperfection of competition, by means of subsidies or by altering the number and size of firms, in order to make price equal to marginal cost in all situations (*op. cit.*, pp. 53–57). T. Scitovsky is more realistic about the problems of product differentiation but also adheres to a marginal costing formula for an "efficient" allocation of resources (*Welfare and Competition*, chaps. 7, 15, 16 and 18). On the other hand, Samuelson does not even mention the problem of applying welfare theories to product differentiation phenomena (*Foundations of Economic Analysis*, chap. 8); nor does Boulding in his survey of welfare economics (*Survey of Contemporary Economics*, Vol. II, chap. 1).

The only healthy scepticism one can detect in the literature is that of Little, who is clear that product differentiation involves "non-marginal" characteristics and that "any general charge of maldistribution of resources on the grounds that competition is not perfect cannot be substantiated from the basis of welfare theory." But he does not develop this theme (*Critique of Welfare Economics*, pp. 259–62).

Chamberlin's strenuous efforts[9] to distinguish "monopolistic" from "imperfect" competition, together with his insistence that indivisibilities are irrelevant, and that the derivation of "imperfection" is in the demand conditions of a good, are more than justified by the seriousness of the errors which have so clearly resulted from ignoring the product differentiation aspects of economic welfare.

VI

The conclusion to be drawn from the particular form of this analysis is that there is within imperfect competition an area, created by the form of consumer demand, which is not amenable to adjustment in the direction of "better" allocative patterns of the economy. Nor would it be desirable to go far with such adjustments were they possible since the presumption is that they must undermine what the consumer considers to be part of his welfare. Marginal cost pricing (or price proportional to marginal cost) as a universal precept is not compatible with product differentiation.[1] The welfare problem presented by product differentiation probably requires a quite different approach. If the presumption is that consumers should have the right to demand a heterogeneous product in the markets then the welfare economist is under an obligation to support policies, such as compulsorily informative advertisement and labeling, approved specifications by consumer institutions, etc., which seek to provide as much guidance as possible for the consumer as to the nature of his choice. On the other hand, it must not be expected that given an increase in information there will be less variety. Conceivably, as consumers become educated to a more precise knowledge of individual variations there could be more variety.[2] This raises the very real problem which unrestricted differentiation of product may create in a free market economy. There seems to be no way open effectively to put before the consumer the choice between a limited variety at low prices and a greater variety at higher prices. By concentrating on the advantages of individual varieties of a product the consumer is likely to overlook the more abstract advantages of general cheap-

9. In the articles already cited, in chap. 9 of *Monopolistic Competition* ("Monopolistic or Imperfect Competition?" this *Journal*, Aug. 1937 and May 1938) and in "Monopolistic Competition Revisited," *Economica*, Nov. 1951.

1. Cf. E. H. Chamberlin, "Product Heterogeneity and Public Policy," *op. cit.*

2. *Ibid.*, on all these points.

ness. And the producer is so organized that it is often to his individual advantage to encourage this tendency with substantial help from the market through oligopolistic and related forces. Consequently there may be "excess capacity" with unrestricted product differentiation.[3] A case could be made, at least for some types of good, for public intervention in order to impose some degree of standardization of product. And certainly manufacturers should be encouraged in this direction by the practices of standards institutions. There is no ignoring the seriousness of the "excess capacity" problem and the general problem of a welfare criterion for diversity; it is crucial for public policy in this field. But it is a problem requiring the somewhat pedestrian accomplishments of common sense and good judgment rather than complicated manipulation of factors and firms.

The line of argument chosen in this article — that of stressing the positive rather than the negative aspects of product differentiation — is simply the working out of a particular problem in order to assess the *appropriateness* of the welfare model for actual conditions. (One could equally well consider whether the positive contributions to welfare of large-scale monopolists and oligopolists do not outweigh the loss in allocative efficiency caused by the failure of marginal cost pricing).[4] A welfare economics should have some orientation towards the *application* of policy if the adjective "welfare" is to have the special significance which it appears to convey. It seems as though the workaday methods of applied price theory — a pragmatic approach which treats each case of "imperfection" of competition on its merits — constitute a more useful apparatus for policy than the too-general models of welfare economics.

ALEX HUNTER.

UNIVERSITY COLLEGE OF NORTH STAFFORDSHIRE,
ENGLAND

3. Chamberlin, *Monopolistic Competition*, pp. 104–9.
4. Cf. J. A. Schumpeter, *Capitalism, Socialism and Democracy*, chaps. 7–8. The British Monopolies Commission appears to believe there is substance in this view; cf. Alex Hunter, "The Monopolies Commission and Economic Welfare," *Manchester School*, Jan. 1955, pp. 31–37.

THE SETTING OF ENTREPRENEURSHIP IN INDIA*

By Andrew F. Brimmer

I. Introduction

The purpose of this paper is to introduce non-Indian readers to the existence in that country of a rather unusual type of industrial organization known as the managing agency system. An attempt will be made to explain its origin and growth and to indicate some of the contributions it has made to economic development. Another objective of the paper is to call the attention of Indian students to an approach to the study and appreciation of this institution which is different from the view most frequently taken. The legal aspects of the managing agency system are usually focused on, and most investigations which the author has seen rapidly become inquiries into, the nature and provisions of Indian company law and the agents' behavior within these limits. The result is that the "illegal acts" and transgressions of specific managing agents constitute the subject matter for discussion.

Our present concern is not primarily with the correctness of this approach, considered by itself, but with the substance of economic activity obscured by preoccupation with the managing agency system as a legal entity. In a wider context, the author would argue that the whole system of company law in India, like the organization of commercial banking, is simply irrelevant. Both of these represent the importation of British institutions into an environment which was alien and inhospitable. However, this broader field is not the one examined here. This writer asserts that legal and economic arrangements developed in the United Kingdom to solve problems arising in the latter were unsuited for India. It is suggested, therefore, that the managing agency system is the Indian answer to the imposition of British institutions as well as a response to the challenge of relative backwardness which characterized the economic scene in India.

II. Nature and Origin of the Managing Agency System

The managing agency system is a type of industrial organization unique to India in which the promotion, finance, and administration of one or more legally separate and presumably independent companies are controlled by a single firm. It is a system of economic and legal relationships which pervades the entire field of economic activity — especially modern industry and trade and commercial agriculture.[1] The system operates within the framework of company law, and both public and private limited joint stock companies as well as partnership undertakings are controlled by managing agency firms. At the very center of the system is the managing agency firm which may take the organizational form of either a partnership, private or public limited company, or a single individual. The formal and informal relationships between this firm and the operating companies controlled by it constitute the managing agency system. The name of the system arises from the so-called "managing agency contract" between the agency firm and each company.

A. Critical Role of the Agency Firm

It is extremely important, if the system is to be understood, that the nature of the managing agency firm and its relation to the controlled company be fully comprehended. It should be noted that the managing agency firm *is the firm* in the sense in which this term is known in institutional economics. If the firm is defined as the institu-

1. P. S. Lokanathan, *Industrial Organization in India* (London: George Allen & Unwin, Ltd., 1933), p. 5.

tional setting in which entrepreneurial decisions are made, it is immediately clear why the managing agency firm should be so designated. Managing agents, the businessmen operating through the managing agency firm, are the real entrepreneurs in India. They have been the ones primarily responsible for the introduction of new products, new methods of production, and new sources of raw materials; they have discovered and exploited new markets and have usually undertaken whatever reorganization Indian industry has experienced.[2] To achieve these ends, the managing agents have generally made use of the joint stock form of organization for the companies launched to undertake actual production and trade. These latter companies should be considered as operating units of the central, decision-making unit — the managing agency firm.

Whatever the legal form which the managing agency firm assumes, it is typically composed of a group of individuals who have either large financial resources (or access to such) or a considerable amount of general business and technical ability, or both. The composition and functioning of the agency firm may vary depending upon whether it is controlled primarily by Indian or British businessmen. (Why this is so is explained below.) The managing agency firm has many instruments by which it maintains control over the operating company. The most important of these are: (1) the management contract, (2) financial practices, and (3) share and voting arrangements. The management contract is an agreement between the agency firm and the operating company which specifies their respective rights and obligations. It is drawn within the framework of the Indian Companies Act, 1913 (amended, 1936). Consequently, the agreement is a legal document and has status in the law courts. Financial holdings of the agency firm in the controlled companies were once the chief means of securing control. In recent years there has been a tendency for agency firms to reduce their holdings of shares issued by operating companies, but they continue to occupy a crucial place in matters of finance. A proven instrument for control exists in the use of shares with special voting rights which are issued to the managing agents only.[3]

Although the operating companies may appear to be independent entities with their own boards of directors, they are usually legal and accounting fictions which serve to give the agents a greater degree of

2. Daniel H. Buchanan, *Development of Capitalistic Enterprise in India* (New York: 1934), p. 145.

3. M. A. Mulky, *The New Capital Issue Market in India* (Bombay: New Book Company), p. 75.

freedom, financial and otherwise. In most cases the board of directors of the operating company is selected by the agency firm.[4] In practically every instance several members of the agency firm are also directors of the controlled company. More often than not, one of these agency directors is chairman or managing director of such a company. While the managing agent presumably functions under the supervision of the board of directors, the latter is frequently nothing more than a fiduciary body which exists to persuade the public to invest and to fulfill legal requirements. A look at the large number of "prestige" directors — "Sirs," "Rajas," "Rao Bhadurs," etc. — on the boards of Indian companies will make this suggestion quite obvious. Thus, the managing agency firm is responsible for practically all decisions made in the companies under its control.

B. British and Indian Agency Firms Contrasted

The above picture of the managing agency system is a generalized one. If British and Indian agency firms are examined separately, some modifications must be made in regard to details. However, the general outlines of the system are not greatly changed. The typical British agency firm is a partnership which continues to combine financial resources and technical and business ability through the selection of its members.[5] Such a firm is likely to have one or two "old family" men who have inherited their position. They may or may not be technically competent. However, they usually are, and if they are not they are unlikely to be in a position to obstruct the efficient operation of the agency firm. In either case, they give continuity to the firm through the maintenance of the family name. In addition to such family representatives, there are always two or three partners selected from the group of senior assistants. The latter officers have usually progressed through the various operating companies under the control of the agency firm and have acquired considerable knowledge of their activities. In most cases, they have begun as junior assistants brought out from England and Scotland in their early twenties. Frequently, they come with the idea of making a career of service with an agency firm and eventually becoming a partner in it. Many succeed.

Each partner in the British agency firm usually rotates as executive officer of one or more of the "departments" within the firm. A

4. Bombay Shareholders' Association, *Memorandum on Directors* (Bombay: 1949), pp. 8–9.

5. Lokanathan, *op. cit.*, and Wilfred Russell, *Indian Summer* (Bombay: 1951), p. 47.

typical British agency firm will have the following departments: accounts and finance, export and import, labor welfare, mills and works, purchasing, research and development, sales and advertising, transportation.   In addition, each partner will serve as director, managing director, or senior officer of one or more of the operating companies.   When the nonfamily partners retire from the agency firm, they are almost never replaced by their sons.   Consequently, the British agency firm is constantly drawing in new blood and new money when the senior assistants become partners.[6]

The typical Indian agency firm differs from its British counterpart in several respects.   However, these differences can be understood only in the context of the traditional organization of Indian business.   For centuries practically all business activity was concentrated in the so-called trading communities.   The latter are remnants of the process of fragmentation through which the older fourfold caste system has gone.   The origins of this system need not detain us here, for it is sufficient to note that occupational stratification was one of its cardinal features.[7]   Within each Hindu trading community the common unit of economic activity was coterminous with the familial organization.   The eldest male, so long as he was not incapacitated, was the chief administrator of the family fortune. Other male members — and sometimes females — participated in the activities of the firm under the direction of the family head and shared (not always equally) in the gains.   As long as the family remained a unit, the firm continued to function, for both tradition and law clearly specified the rights of succession.[8]   This was, of course, a great element of stability, but it may not have fostered maximum efficiency.   Finally, there existed between the family-firm and the trading community of which it was a member an informal relationship symbolized by a very strong sense of responsibility for the well-being of one's community fellows and an overt preference for dealing with them.

Of course, it is not necessary to dwell upon the fact that the traditional bonds and conditions discussed above have been weakened considerably.   The trading communities are no longer the sole Indian participants in business activity.   The joint-family has lost much of

6. Russell, *op. cit.*, p. 58.
7. D. R. Gadgil, "Notes on the Rise of the Business Communities in India" (New York: Institute of Pacific Relations, 1951), pp. v–vi.
8. G. S. Ghurye, *Caste and Race in India* (London: 1932), pp. 28–29.   And Sir Edward Blunt, "The Economic Aspects of the Caste System," in R. Mukerjee (ed.), *Economic Problems of Modern India* (London: 1939), p. 64.

its former hold on the individual members, and the family-firm is no longer the only crucible of family fortunes. The overt preferences for caste fellows is not so frequently met as it once was. Nevertheless, the legacy of the past has a reality all its own and serves to shape the mode of operation of the Indian managing agency firms.

The managing agency form of organization was adopted quite widely by the Indian merchants when they began to operate in the industrial field. But unlike the British agency firms, the Indian agency houses are primarily financial in character.[9] Apart from a few exceptions,[1] the Indian agency firm is typically a simple extension of the older family-firm relationships. This fact is reflected in the composition of the firm, its sources of finance and the way in which it operates the companies under its control. While one encounters Indian agency firms organized in the form of private limited companies, the usual form is a partnership. Within the partnership, the members are most frequently related by blood or marriage, and in practically every case, they are — at the very least — members of the same caste. The sons of the partners, irrespective of ability — tend to follow the fathers in unbroken succession.[2] Despite the relatively large size of Indian families, this means that the field from which senior members of the firm are recruited is extremely limited. In addition, the Indian agency firm tends to look for junior assistants and other staff members first among the members of the agents' family and secondly among their community or subcaste. Only very infrequently do they go beyond these bounds to obtain persons for staff and administrative posts.[3]

Since members of the Indian agency firms, in general, have little technical competence, they occupy their position chiefly on the basis of the family funds invested therein. These funds have frequently been accumulated through trade and constitute the major contribution which the Indian managing agents make to the agency firm as well as to the operating company or companies under the latter's control. Another way in which the financial character of the Indian firms is revealed is the transfer of managing agency rights in an operating company from one agency firm to another. The receiving agency firm almost invariably has greater financial resources than

9. Lokanathan, *op. cit.*, pp. 315–19.
1. The most widely known of these exceptions are the House of Tata, Birla Bros., Khatau & Sons, and Wadia and Sons, Ltd.
2. Gadgil, *op. cit.*, p. vii.
3. Leslie Sawhny, "Management and Training in Industry," *Commerce and Industry* (Madras: February 14, 1954).

the vending firm and only occasionally is the former superior to the latter in the way of industrial experience.[4]

With a few notable exceptions,[5] the Indian agency firms seem to administer the operating companies under their control with a view to obtaining the maximum amount of profit in the shortest possible time. There appears to be little attention given to the maintenance of plant and equipment or to the long-run planning and development of markets and raw materials. As indicated above, there is practically no effort made to recruit and train a staff on the basis of objective personnel criteria. Though the technical staff of most Indian agency firms must be employed, questions of tenure and job security apparently are seldom, if ever, discussed. When Indian firms adopt accounting procedures (as required by law for the public limited companies under the agents' control), they often do so in order to conceal rather than explain the operations of the company.[6]

In general, most of the British agency firms approximate in structure and functioning the desirable aspects of the managing agency system. With few exceptions Indian firms seem to embody many of the undesirable features. In reaching this conclusion, it is necessary, of course, to distinguish between the desirable and undesirable features of the managing agency system and those practices followed by specific agency firms. The contributions which the British agency firms have made to the development of Indian industry are contributions made by seasoned entrepreneurs operating through the agency firm. Likewise, the detractions made by Indian agency firms have been made by businessmen still in the process of maturing. Consequently, the worthiness of the system must be judged in terms of what use various business groups have made of it. In the hands of the British and the more advanced Indians, the system has been a powerful instrument of economic development; in the hands of the more irresponsible Indian businessmen, it has been an equally powerful instrument for manipulating enterprises to the advantage of controlling families but usually to the disadvantage of the country as a whole.

C. Origin and Development of the Managing Agency System

The exact origins of the managing agency system are still disputed. However, most writers on the subject seem to agree that its

4. Bombay Shareholders' Association, *Memorandum on Managing Agents* (Bombay: 1949), pp. 1–31.

5. See the agency houses listed in note 1, p. 558.

6. S. D. Mehta, "Some Aspects of the Managing Agency System," *Journal of the University of Bombay*, Jan. 1953.

roots go deep into the social and economic conditions prevailing in India at the time the East India Company lost its monopoly of trade. The evidence indicates that the system evolved out of attempts to overcome two limitations on the appropriation of business opportunities then prevailing in India:[7] (1) a shortage of entrepreneurial ability and (2) a shortage of venture capital. The continuing presence of these limitations seems to go a long way in explaining the continued existence of the system.

During the early nineteenth century, India offered a number of opportunities to energetic businessmen. Its large population, despite the low standard of living, presented a relatively large demand for consumer goods — especially cotton textiles. Although the latter were widely produced in India, the methods of production were backward when compared with the power-driven mechanical techniques which were then revolutionizing industry in England. Also, India's large capacity to produce industrial raw materials was as yet underdeveloped. At the same time, the supply of business leadership was small. There were very few indigenous Indian businessmen who were able to move from the traditional trading and financing in local markets to the new forms of production. To profit by the opportunities then available, required a knowledge of foreign markets and sources of supply. It also required a mastery of new forms of organization and widespread connections in the world economy then in the process of emerging. The field was open in India, and the British general merchants were on the scene.[8]

There is some evidence which suggests that several British firms were trading in Bengal in the late 1790's with arrangements similar to those which characterize the modern agency firm.[9] In general, it seems that former servants of the East India Company set out as general merchants on their own account. As general traders and factors, they acquired a detailed knowledge of local markets and ways to exploit them. Given their connections with firms in England, America, and other parts of the world, this new knowledge of Indian conditions placed these merchants in a position to appropriate new opportunities as they emerged.

In addition to the relative shortage of entrepreneurial ability, there was also a drastic shortage of venture capital. Despite the numerous fortunes accumulated by Indians in trading and agricultural activity, the supply of funds to finance the projects of the British

7. Lokanathan, *op. cit.*, pp. 15–16.
8. William W. Hunter, *The Indian Empire* (London: 1893), pp. 658–59.
9. *Ibid.*

general merchants came primarily from the merchants and a few of their friends and associates.[1]   The usual procedure followed by the agents in the launching of new enterprises was as follows: As promoters, the agents would make the necessary investigations concerning cost and availability of labor, raw materials and other inputs and the nature and extent of markets.   The agents would also organize and register the company as a legal entity, selecting in the process the first board of directors.   Initially, the agents would issue only a small part of the authorized capital of the new company and — with their friends and associates — would take up most of the issued shares. If debentures were floated, the agents would also hold these.   Once the enterprise had proved itself, the agents would issue most of the remainder of the authorized capital — usually preference shares — and invite the public to subscribe to the securities.   At the same time, the agents would very likely throw some portion of the shares originally held by themselves onto the market.   In this way, the agents supplied capital to the new enterprise in the most critical time.[2] Once he had regained his initial investment — frequently with a considerable profit — the agent was able to repeat the process with a new enterprise.   As will be more fully explained below, the agent was also able to shift funds from one enterprise to another through the medium of the agency firm.

Although the conditions discussed above may explain why the British merchants developed the managing agency system, the question of why Indian businessmen adopted it must be answered.   The first clue, of course, must be sought in the same conditions which made the British develop the system.   The supply of Indian entrepreneurs was equally as limited as that of the British.   The more enterprising of the Indian traders (notably Parsis and Gujeratis) turned to trading in cotton and similar goods soon after the British came to India.   By the middle of the nineteenth century, the Parsis, who had forged the strongest connections with the British whom they served initially as middlemen, were able to branch out into the manufacture of cotton textiles.[3]   The latter industry has been the basis of most of the accumulations which have gone into other industries in which Indians have played a major part.   The agency system proved a highly useful framework within which business ability could be put to the best advantage.   Like the British, Indian businessmen found

1. Russell, *op. cit.*, p. 63.
2. Lokanathan, *op. cit.*, pp. 14–27.
3. Buchanan, *op. cit.*, pp. 270–71 and Gadgil, *op. cit.*, vi–vii.

that the agency system could be used to overcome the handicap of capital shortage.

Another factor contributing to the widespread use of the system by Indians lay in the nature and functioning of the banking system in India.[4] The loan criteria used by the Imperial Bank of India, and other leading joint stock banks before it, gave to the managing agents a crucial role in obtaining bank credit. Like the English banks on which they were patterned, Indian joint stock banks limited their loans to the supply of working capital, accepting the hypothecation of goods as security. However, the banks argued that goods were likely to vary greatly in price — and price instability was especially true of cotton and piece goods in which Indians traded to a large extent — and it was, therefore, necessary to have someone on whom the bank could rely to make good the loan. This someone was clearly the managing agent; for while Indian joint stock companies showed a high propensity to go bankrupt, the agency firm displayed a much higher survival rate. As a result, the practice grew up of having two signatures on notes of indebtedness, one a director of the company and the other the managing agent.[5] Thus, in order for a company to obtain even the limited amount of credit the banks were willing to advance, it was mandatory for the company to be under the control of a managing agent.

Finally, as indicated above, the agency system possessed much flexibility from the point of view of a small group of businessmen with a limited amount of capital but looking to the development of industrial and commercial enterprises. At the same time, the system was equally as flexible in the hands of persons who were primarily interested in manipulating finances among various companies for the sole purpose of enriching the managing agents. An agency firm was able to transfer funds from one company to another regardless of the soundness of the receiving unit. A prosperous undertaking could be made the basis for raising funds in the capital market for use by companies whose own credit standing would not warrant confidence.[6] Because of the concentration in the hands of the managing agents of purchasing and marketing arrangements of various controlled companies, the best interests of any individual company were not always served. This ability to conceal the actual status of each operating company behind the multiplicity of interlocking balance sheets also

4. S. K. Muranjan, *Modern Banking in India* (Bombay: 1952), pp. 289–321.

5. D. R. Samant and M. A. Mulky, *Organization and Finance of Industries in India* (Calcutta: 1937), p. 45.

6. Bombay Shareholders' Association, *Managing Agents*, pp. 41–74.

gave the agents a considerable amount of leeway if they chose to utilize it in escaping taxation of earnings. The evidence suggests that many Indian agency firms selected the agency system as much for these predatory activities as for the more socially desirable ones described above.[7]

Despite the transgressions which the managing agency system allows to businessmen who wish to make them, one cannot avoid concluding that the system has gone a long way in reducing the limitations and disadvantages which enterprising businessmen had to face in their attempts to answer the challenge of industrial backwardness in India. For one who wishes to explore them, the numerous problems created by the separation of ownership and control in joint stock enterprises are multiplied many times in the agency system. But this writer argues that if one focuses not on the legal aspects of the agency system but on its economic characteristics, the managing agency firm comes into full view as the heart of industrial organization in India. Consequently, if this organization is to be fully appreciated, the numerous, legally distinct, producing companies must be treated simply as operating arms of the managing agency firm.

III. The Scope of the Managing Agency System

No matter how extensively one reads in the various standard works on Indian economics, it is not possible to obtain a clear view of the place the managing agency system plays in that country's economy. A comprehensive, quantitative appraisal of its scope is yet to be made. It is obvious, however, that such a task will be almost impossible to carry out. In the first place, as explained above, the activities of managing agency firms — while the substance of Indian enterprise — are generally hidden behind the shadow of joint stock and private companies through which the agents operate. The operating companies, when publicly registered, constitute the visible superstructure: beneath them, in the byways of partnership and private arrangements, are the crucial mechanisms of business activity encompassed by the managing agency firm. Almost no public records exist to illuminate these.

Secondly, much of general economic activity is not carried on through the instrument of publicly registered joint stock companies. Under the Indian Companies Act, 1913 (amended, 1936), any group containing not more than fifty members can register itself with the proper authorities as a private company with limited liability. It can

7. *Ibid.*

issue stock of a specified nature, and so long as this stock is privately held, it can engage in any kind of business except scheduled banking and insurance.[8] It is in this form that a significant portion of Indian business is carried on.

A. Industrial Distribution of the Agency System

Nevertheless, the need for such statistical knowledge is obvious. Otherwise, discussion of the managing agency system must continue to have an air of unreality about it. In an effort to cast some light on this area, the author undertook an exploratory investigation using data from one of the largest *Yearbooks*[9] which contains information relating to companies registered on the various Indian stock exchanges. Of the 1,427 companies for which complete information was given, 94 were banks, 70 insurance companies and 20 managing agency firms. These 184 companies were set aside, the banks and insurance companies for the reason explained in note 8 above, and the managing agency firms because they were not likely to have other managing agency firms directing them. Thus, there were 1,243 companies studied. All of these could have been legally controlled by managing agency firms. The number with managing agency contracts was 1,064 — leaving only 179 companies without managing agents. This meant that managing agents dominated more than 85 per cent of the companies listed on the stock exchanges. While it is not possible to give a close estimate of the proportion of total production handled by these firms, other data suggest that the ratio is very high in the industrial sector.[1]

The 1,064 companies were classified under seventeen industries. There were three industries, coal, shipping and cement, in which all the companies found were controlled by managing agents. In four others — jute, cotton, railways and North India Tea estates — more than 90 per cent of the companies were under the direction of managing agents. These seven industries had 523 or 42 per cent of the companies analyzed. With the possible exception of railways (most of which are owned by the government) and cotton textiles (in which handicraft and small-scale units play a large part), the publicly

8. Scheduled banks are those with deposits of Rs. 500,000 and are included within the Reserve Bank System. Both scheduled banks and insurance companies must be publicly registered joint stock companies. Neither can enter into a managing agency contract with another firm or individual.

9. *Investors' Encyclopedia* (Madras: Kothari and Sons), 16th Edition, 1951–1952.

1. Mehta, *op. cit.*

registered joint stock companies are responsible for the bulk of the output in these industries. These are among the most fully developed industries in India.

Four other industries in which managing agents controlled more than 80 per cent of the producing units are engineering, electric power, vegetable oils and planting (mostly coffee, rubber and nuts). Out of 310 companies in these industries, 265 (85 per cent) had managing agents. It should be noted that the engineering industry contains such crucial firms as those producing iron and steel, industrial machinery and agricultural implements. Actually, industries in which 70 per cent of the firms are controlled by managing agents probably could be classified as being under the latter's domination. Sugar, paper manufacturing and mining and chemicals are in this category. In conclusion, one can say with a high degree of confidence that the managing agency system permeates the entire field of modern industry in India in which private enterprise is most active.

B. Type and Size of Agency Firms

The exploratory study also revealed another interesting feature of the managing agency system. The 1,427 companies analyzed were under the direction of 844 management units. These latter consisted of 524 managing agency firms and 320 individuals. With respect to the agency firms, it was found that the private limited company is the most popular organizational framework. About 66 per cent (346) of the agency firms fall into this category. The number of partnerships (156) is less than half as large. Public limited companies were 22 in number and only 5 per cent of the total management firms. Two of the managing agency firms with public registration were actually operating companies which also held managing agency contracts with other operating companies. There were 363 companies with individual managers. This number includes the 184 companies which were either banks, insurance companies or managing agents. It follows then that only 179 of these companies could have had managing agents. Further, the bulk of these 363 companies (82 per cent) was managed by an individual with only one company under his direction. Even so, one should not conclude that such an individual was not a hired servant of some managing agency firm. For example, it is reported that some of the large banks are owned or controlled by several of the leading managing agents. (Tata is said to control the Central Bank of India, Birla, the United Bank and Dalmia, the Bharat Bank). The 23 individuals who managed the

remaining 66 companies seem to have between two and three companies each under their direction.

What is probably one of the most interesting aspects of the managing agency system was thrown into focus by the study. It was discovered that 77 per cent of the managing agency firms (404) controlled only one company each. On the other hand, the remaining 120 firms controlled 660 operating companies — 62 per cent of the total number of companies controlled by agents. From these statistics, it is obvious that the managing agency system is not composed of firms possessing comparable economic power. On the contrary, there appear to be scattered groupings of very large firms surrounded by an extensive field of small units. What is the reason for the existence of the one-company agencies?

The author confesses that he has no satisfactory explanation. However, certain aspects of the matter may point toward an answer. The first thing to note is that every one of the 404 one-company agencies is Indian. Secondly, the bulk of them is found in cotton textiles and among such miscellaneous industries as soap-making, biscuit and bobbin manufacture, coir, glass, woodworking, etc. In this section of the cotton textile industry as well as in the miscellaneous industries, the amount of capital invested in each operating unit is quite small. There is almost no public participation despite the adoption of the joint stock form of organization.[2] The incentives to use the managing agency system were discussed in Section II. Of the 404 one-company agencies, 255 represented that many *separate* businessmen or associated groups. These 255 perhaps can be explained by the conditions mentioned in the above discussion. The data indicate, however, that there are 28 individual businessmen in control of the remaining 149 one-company agencies. This is an average of about *five one-company agencies per individual.* Why do these businessmen follow the practice of setting up a new agency firm when they decide to undertake an additional producing activity? It has been pointed out that some of the largest British and Indian agency firms control operating companies in such diverse fields as cement, cotton textiles, jute, sugar, shipping, planting as well as trade. Why, then, this effort at separateness among some Indian businessmen?

The reason for this behavior probably lies in the fact that most Indian businessmen launch their business careers as traders and continue to rely on trade as the chief source of income.[3] If such a businessman were to concentrate all of his assets in one agency firm

2. Samant and Mulky, *op. cit.*, pp. 224–49.
3. Gadgil, *op. cit.*, pp. 29–32.

and engage in both trade — a relatively safe activity — and experimental manufacturing, he would expose his entire fortune to the vicissitudes of a field in which he possessed little competence.   If the manufacturing ventures employed borrowed funds (and most would have to borrow working capital even when the businessmen as owners and agents supplied the fixed investment), the managing agents would have to guarantee the loan personally.   Under these conditions, it is safer to launch a new agency firm for each project than to group several unproven projects under the same firm.   The financial losses arising from the risk of failure of a single operating enterprise are restricted to that enterprise and its immediate managing agency firm; the gains from the same can be appropriated by the businessman who controls all of them.

From the above discussion, it is clear that the managing agency system has been an important instrument for reducing the limitations imposed on industrial growth by shortages of capital and business ability.   It has also been a powerful lever in the hands of those who wish to manipulate investments and investors so as to improve their own positions rather than that of the country as a whole.   The data presented show that the system is employed throughout the industrial sector of the Indian economy, and its influence extends even further.

IV. Relations Between the Agency Firm and Operating Companies

The scarcity of material for studying the relations between the managing agency firm and the companies controlled by it is extremely difficult to overcome.   It is practically impossible for one not connected with the firm to get inside to observe its organization and functioning.   Such a position is required, however, before an account of the decision-making process can be given.   This author hastens to admit that he was not so fortunate as to study any agency firm from that vantage point.   The writer did have the opportunity of meeting representatives from several agency firms who were kind enough to answer most of his questions.   In addition, officers of the Bombay Shareholders' Association, who concern themselves primarily with the affairs of companies in which the Association's members have invested, were quite helpful in providing material from their files.   These data, combined with information from the stock exchange yearbooks, made it possible to draw several of the major outlines of the relations between the agency firms and the operating companies.

Available materials throw light on four areas of agency-company interrelations. These concern the nature of the managing agency contract and the rights and obligations it specifies, financial relations between managing agents and controlled companies, the use of special shares to maintain control, and the position and authority of company directors.

A. The Managing Agency Contract

It is not an easy task to obtain copies of managing agency agreements. The author, through the facilities of the Bombay Shareholders' Association, was able to examine two dozen of these contracts.[4] These were said to be representative of the general contents of such agreements. The more common aspects are the following:

Each agreement specifies the parties to it (agency firm and operating company) and sets forth the period of time during which it is to be effective. The conditions of termination and renewal are also indicated. The average duration of the twenty-four contracts was twenty years. Five were made for a period of thirty years. Although this group of contracts contained none with a life longer than thirty years, the author was informed that several companies investigated by the Bombay Shareholders' Association had contracts which covered the entire life of the company. The Government of India is currently considering a proposal which would limit the length of managing agency agreements to fifteen years.

One of the most important provisions of the agreement specifies the amount of remuneration the managing agent will receive and the methods by which it is to be computed. A typical provision sets compensation in the form of a commission of 10 per cent of annual net profits of the company. Other provisions usually specify whether the managing agents are to share in profits by the operating company from the sale of assets or capital gains. In most instances, companies pay the managing agents a monthly allowance to defray the cost of office administration on behalf of the operating company. Among the twenty-four contracts examined, eighteen provided for office allowances ranging from Rs. 1,000 to Rs. 10,000 per month, the average being Rs. 2,500.

Although most agency agreements pledge the agents not to undertake business on their own account in direct competition with the managed company, few prevent the agents from acting as manag-

4. As a condition for obtaining access to these contracts, the author had to agree not to identify the companies concerned.

ing agents for other companies engaged in the same type of activity. That this situation frequently leads to conflicting interests between agents and the different companies under their control has been commented on many times. However, if one takes the view presented in this paper — that the managing agency firm is the crucial decision-making unit and the companies under its control are simply operating arms — this conflict will lose much of its apparent significance.

A study of the twenty-four contracts mentioned above, shows the wide scope the agency firm has in the affairs of the controlled company. Subject to the control of the directors of the company (extremely weak), the managing agents have the general conduct and management of the business and affairs of the company. The agents have the power to enter into all contracts (except the one with themselves!!) on behalf of the company. They have the authority to initiate and abandon on their own decision all legal proceedings involving the company. The agents are solely responsible for the purchase, maintenance and sale of all raw materials, and with the sanction of the board of directors they decide all expenditures on fixed investments. Likewise, the agents have complete control of the financial affairs of the company. It is they who receive all money payable to the company from all sources; it is they who pay all bills. Their signature appears on all financial instruments, and they alone buy and sell government bonds on behalf of the company. All employees of the company — professional, clerical, skilled and unskilled — are engaged and dismissed by the managing agents. These employees are placed on the payroll of the company, but it is the managing agents who decide the salaries, wages and other benefits they receive.

B. FINANCIAL RELATIONS

The financial holdings of the managing agency firm in the operating companies were once considered to be the primary basis of the former's control over the latter. This is no longer true. Although the extent of holdings will vary, almost every agency firm will hold some portion of the capital of the companies under its control. Different writers have given percentages for the agency firm's share which show considerable range. These have varied from as low as 1 per cent to as much as 75 per cent.[5] Of course, it is practically impossible to obtain reliable information on this question on any significant scale.

5. See the Appendices to chap. I of P. S. Lokanathan, *Industrial Organization in India* (London: 1933).

The author was able to obtain information which may cast some light on the subject. The statistics in Table I relate to the holdings of a Bombay managing agency firm in seven operating companies under its control, its holdings in another managing agency firm, and the extent of investment of the first agency firm in one of the operating companies controlled by the latter.

Because of the lack of comparative data, the author is not in a position to appraise the representativeness of these figures. Nevertheless, the table is interesting in itself, for it gives some indication of the variety of activity in which a managing agency firm engages.

A more important aspect of industrial finance in India under the managing agency system is shown in Table II. The meaning of the column headings is as follows: Column (2) includes the funds raised in the capital market on the account of one company under the managing agency firm listed in column (1) but subsequently transferred to another company under the same agency firm. Column (3) shows the amount of loans and advances made by one operating company to another, both being under the control of the same agency firm. Likewise, column (4) shows intercompany investments by companies under the same agency firm. Book debts in column (5) are those which reflect the net value of goods and services supplied on credit to each other by various operating companies. Column (6) is the net amount loaned to the agency firm by the operating companies. The number of companies involved in each group varies widely, from more than fifty on the part of the Dalmia and Birla Groups to less than five for the Narang agency firm. The data have been compiled from the source indicated in the table.

The financial advantages of the managing agency system when looked at from the point of view of the agency firm are quite obvious when one realizes that the behavior pattern indicated in the table is followed quite generally by a considerable number of managing agency firms, British and Indian. With the agency firm in a position to transfer funds from one operating company to another at will, it is clear that an agent can nourish or strangle any company under his control to any extent he desires, his power to do so being limited only by the total amount of funds available to all the companies. The ability of the agency firm to carry out such financial maneuvers is often criticized by stockholders, but the latter are helpless to do anything about it. They must rely on either their vote or the authority of the directors of the company to prevent such practices. As shown below, the directors are mostly straw-men shield-

TABLE I

KILLICK INDUSTRIES, LTD., INVESTMENTS IN OPERATING COMPANIES (RS. IN THOUSANDS)

Name of Operating Company	Capital of Company			Agency Holding			Ratio of Agency Holdings to Total Equity		
	Total	Equity	Debt	Total	Equity	Debt	Total	Equity	Debt
Ahmedabad Electric	47,466	34,966	12,500	993	793	200	2.1	2.3	1.6
Bombay Suburban Electric	9,373	9,373	—	500	500	—	5.3	5.3	—
C. P. Railway	9,400	9,400	—	n.a.	n.a.	n.a.	—	—	—
Hingir Rampur (Metal Mining)	1,200	1,200	—	237	237	—	19.7	19.7	—
Kohinoor Mills (Cotton)	10,000	10,000	—	449	449	—	1.5	1.5	—
Shrivjapur Syndicate (Mica Mining)	525	525	—	10	10	—	1.9	1.9	—
Surat Electric	2,500	2,500	—	103	103	—	4.1	4.1	—
Total	80,464	67,964	12,500	—	—	—	—	—	—
Cement Agencies, Ltd.	800	800	—	240	240	—	30.0	30.0	—
Associated Cement Co.	88,038	88,038	—	569	569	—	0.6	0.6	—

Source: *Commerce* (Bombay: Nov. 1948).

n.a. = not available.

TABLE II

Intercompany Financing Under Managing Agency Firms, 1947 (Rs. in Thousands)

(1) Managing Agency Firm	(2) Transfer of Capital, Market Borrowings	(3) Net Intercompany Loans and Advances	(4) Intercompany Investments	(5) Book Debts	(6) Net Loans to Managing Agents
1. Dalmia Group	16,976	13,991	34,463	12,893	—
2. Walchand Hirachand Group	1,297	5,800	12,958	—	—
3. Narang Group	546	—	3,800	—	—
4. Birla Group	3,000	2,400	10,858	—	—
5. Kamani Group	1,000	704	—	1,200	728
6. Cassamally Munjee Group	—	—	—	—	237
7. A. V. Thomas Group	—	—	507	—	17
8. Andhra Engineering Group	—	—	—	—	225
9. Modi Group	—	2,760	—	—	—
10. Hukumchand Group	487	—	—	—	1,762
11. Karamchand Thaper Group	1,250	1,729	150	—	1,414

Source: Bombay Shareholders' Association, *Memorandum On Managing Agents* (Bombay: 1949), pp. 41–74.

ing the managing agents who can usually muster enough votes through their own holdings and proxies to dominate any shareholders' meeting.

C. Control of Voting Rights by the Agency Firm

Another facet of agency-company relations is the tendency of the promoters of the operating company to keep control in their hands by the issuance of deferred and other types of shares with special voting rights. (These shares are also issued at times for the agency firm itself if it adopts the form of a limited liability company.) Such deferred shares are usually of low denomination with disproportionate rights as to voting, dividends and distribution of assets on winding up as compared with the other high denomination shares. Invariably, the deferred shares are allotted to the managing agents and their associates. Thus, with a small total investment, the managing agents can control the firm, for these shares always have the same voting power as ordinary shares. They do not normally receive dividends until the common shares have received a specified percentage — often between 5 and 8 per cent. Then, however, these securities almost always participate *pari passu* with the common shares, which means that the rate of return on deferred shares is extremely high. The author was able to examine the capital structure of twenty managing agency firms which are public limited companies. Among these firms, three had issued deferred shares. In two cases, the firm had raised 25 per cent of the total capital from deferred shares, but these shares entitled the managing agents who subscribed to them to 50 per cent of the total voting stock. In the other case, deferred shares accounted for only 5 per cent of the capital raised but had 55 per cent of the voting power. The data relating to the operating companies controlled by the agents show that about 60 per cent of the former have issued deferred shares. However, it was not possible to discover who owns them. These examples seem to indicate that the use of special voting shares is resorted to by managing agents as one of the chief means of maintaining control of the operating companies.

D. Role of Boards of Directors

The author heard much discussion in India about revitalizing the boards of directors of joint stock companies by turning over to them the affairs of these companies. Such steps, it was argued, would lead to a closing of the breach between ownership and control

and removal of the managing agents as the dominant force in indus-
trial organization. This author suggests that the legal abolition of
the managing agents would not result in removal of the agents as
controlling factors. Aside from the contractual relations between
the agency firm and the company as defined in the agreement, the
managing agents occupy their position because of the business ability
of the members and employees of the agency firm. Destruction of
the agreement would leave the latter very much intact.

In this area, as with other aspects of the problem, it is difficult
to obtain data with which to appraise the role of the managing agents
in their relations with the controlled company. In an effort to remedy
this situation, the author analyzed data relating to the twenty man-
aging agency firms with public registration mentioned above. Of
course, the generality of these data is limited by the preponderance
of British agency firms among the twenty. The picture for an equal
number of Indian firms would probably change some of the details.

Among the twenty agency firms were four controlled by Indians
and sixteen by British. There were in these firms 120 directors:
77 British and 43 Indian. Of these 120 directors, 73 held director-
ships in the 208 operating companies controlled by the agency firms.
The 208 controlled companies provided 968 directorships. Although
the 73 agency directors were only 15 per cent of the 488 men serving
on the boards of the 208 companies, they held 32 per cent of the
directorships in the latter.

The dominant position of the agents is emphasized more strongly
when one isolates the "working" directors from those found on the
boards for "prestige" purposes. In India, the latter are mainly titled
persons who occupy their seats through the invitation of the manag-
ing agents. Among the 488 men holding directorships in the 208
companies were 129 with titles. Nineteen of the latter were also on
the boards of managing agency firms. All of the remaining 110 titled
directors found only on the boards of the operating companies were
Indians. Their titles included 23 "Sirs" and 87 "Rajas," "Rao
Bhadurs," etc. These 110 purely prestige directors held 269 director-
ships, while 19 agency directors with titles held another 57. If the
agency directors and the prestige directors are combined on the basis
of a common interest, they number 183 persons with 581 director-
ships. This is 38 per cent of the total number of directors with 60
per cent of the total directorships in the 208 operating companies.
This means that the 305 directors not connected with the managing
agency firm in either a working or prestige capacity held only 387

directorships. Thus, managing agency directors held an average of 4.7 directorships in the controlled companies and prestige directors held an average of 2.5 each. On the other hand, nonagency and non-prestige directors held only 1.1 directorships per man.

The above statistics, of course, do not answer the crucial question concerning the distribution of authority among agency and non-agency directors on the various boards of the operating companies. Yet, they suggest that the agency directors occupy strategic positions within the operating companies and are in a position to influence greatly the decisions made in those companies. Consequently, this author is prepared to argue that in nearly all operating companies in India, British and Indian, the role of directors performing inde-pendently of managing agents — or in opposition to them — is most infrequently met with. The managing agency firm dominates the board of directors in the same way that it dominates all other aspects of enterprise in the modern sector of the Indian economy.

V. Conclusion

The conclusions reached by the author in this paper have been set forth at the end of each section. These will simply be summarized here for convenience. It is argued that the managing agency system of industrial organization is the result of efforts by British and Indian entrepreneurs to overcome the limitations imposed by a shortage in India of venture capital and business ability. The managing agency firm must be considered the heart of the system, for the legally sepa-rate joint stock companies in which the producing and trading activi-ties of managing agents are carried on are, in fact, simply operating arms of the key unit — the decision-making agency firm. Although the data supporting this conclusion are not very voluminous and come from varied sources, the author thinks the information is suffi-cient to outline the major contours and behavior patterns within the system.

One area of current discussion of the managing agency system has not been touched on here. That is the nature of criticism which is frequently leveled at it by various persons — many of them highly-placed in government, law, journalism and economics. Most of these censures are directed at the manipulations and transgressions indicated in the body of the paper. Some of the proposed reforms strike at the entire system even though the evidence used to show why they are necessary relate only to the shortcomings of specific managing agents. Whereas the acts of irresponsible agents receive

the most publicity, the achievements of those who use the system as an instrument of much-needed development go unheralded.

The whole question of the managing agency system was studied recently by a company law committee appointed by the Government of India. The lengthy report of the committee boiled down to a set of specific recommendations to remove some of the more obvious opportunities for wrongdoing by agents. These were mainly technical recommendations and need not be discussed here. The Central Government still has the subject under consideration, and will probably legislate a new company law in the near future. Nevertheless, it is doubtful whether the fundamental character of the managing agency system will be changed. For this reason, the author thinks the system is an institution worthy of further study by those interested in economic development in India.

<div align="right">ANDREW F. BRIMMER.</div>

HARVARD UNIVERSITY

ALFRED MARSHALL ON ECONOMIC HISTORY AND HISTORICAL DEVELOPMENT*

By Bruce Glassburner

Alfred Marshall's work is commonly classified as "micro-statics" in modern terminology. He is best known as the synthesist of several strains of nineteenth century thinking, particularly in the fields of cost and utility theory. This emphasis is largely due to the fact that the part of Marshall which has been most widely read and discussed is the "theoretical heart" of his *Principles of Economics*, i.e., Book III through Book VI. That writing includes his theory of the consumer (Book III), theory of production and of business organization (Book IV), theory of market equilibrium (Book V), and his theory of functional distribution (Book VI). In all of this writing, Marshall dealt with the individual, or with small units, and he dealt with them, for the most part, in terms of equilibrium positions, rather than in terms of process.

Marshall's expressed attitude toward the *Principles*, however, was that it constituted only a first approximation. His early intention was to follow the *Principles* with at least one more volume which would be an extension of the first volume and which would be, in some sense, a closer approximation to economic reality. It is therefore appropriate to regard the *Principles* as being what Marshall labeled it, an "Introductory Volume," simplified to make it pedagogically useful. Perhaps the most important aspect of that simplification was the use of the static approach to economic analysis. For Marshall regarded economic life, in reality, as process, and proper analysis as necessarily describing change. However, he argued that because of the evolutionary nature of social change, static pictures of that action which could be regarded as "normal," would not do any violence to fact which would be serious at the elementary level. Thus, he chose *Natura non facit saltum* as the motto of that first volume. It would be for later, more advanced, works to explain the

way in which normality itself could change, and to provide some kind of framework for explanation of "non-normal" activity. Part of the type for the second volume of the *Principles* was set up as early as 1904 (some kind of record was established when it remained set up from then until finally used as part of the text of *Industry and Trade*, published in 1919).[1] *Money, Credit and Commerce*, which appeared in 1923, was to extend the analysis still farther. Marshall wrote that these three volumes were "designed to deal with all the chief problems of economics, so far as the writer's power extends."[2] Presumably, then, with their publication, the advanced as well as the elementary treatment had been presented, and surely treatment of "the main concern of economics," which is "with human beings who are impelled for good and evil, to change and progress,"[3] is to be found in the three volumes. Marshall had entertained hopes that the analysis might be extended still further. A fourth volume was projected, which, Keynes tells us,[4] he made continued effort to complete virtually until the time of his death. The title of that final work was to be *Progress: its Economic Conditions*. His cautiousness, ill health, and time spent in public service slowed his pace and prevented its publication. It is possible that that work might have altered the total perspective of Marshall's contribution to economic thought somewhat, but it seems probable that the three published volumes represent quite well the main thoughts of the approximately fifty years that Marshall devoted to economics.

Marshall's theory of process is sketchy and obscure. Considering his well-known careful thought processes and meticulous effort at logical completeness, the question arises as to whether or not he planned (perhaps for that fourth volume) a detailed, consistent, dynamic model. It appears, however, that Marshall had no intention of doing more in this area than to pull some loose ends together. There is no indication in his formal writings, letters or elsewhere that he intended to develop such a model: on the contrary, there are repeated suggestions that such efforts are inappropriate.

Examination of Marshall's writing seems to indicate that *Industry and Trade* is the work mainly devoted to treatment of process,

1. Alfred Marshall, *Industry and Trade* (2d ed.), p. vi.
2. Alfred Marshall, *Principles of Economics* (8th ed.), p. xi. All subsequent citations will be to this edition unless otherwise specified. Hereafter it will be cited as *Principles*.
3. *Ibid.*, p. xv.
4. J. M. Keynes, "Alfred Marshall, 1842–1924," *Memorials of Alfred Marshall*, ed. A. C. Pigou, p. 65. Hereafter this volume will be cited as *Memorials*.

although in neither of the second two volumes is the "micro-static" analysis of the *Principles* reverted to, except in summary fashion. The bulk of the writing in both *Industry and Trade* and in *Money, Credit and Commerce* is historical, descriptive, and aggregative. There is nowhere an attempt to expound a "theory of development" of the type offered by Professor Schumpeter. Nor is there any effort at dynamic process analysis of the modern mathematical sort *à la* Samuelson. Marshall's attitude toward the nature of economic development was such as to make model-building of this kind inappropriate if not impossible, as will be shown in the discussion following.

Marshall's central idea on dynamic economic problems is contained in his statement that "The Mecca of the economist is economic biology rather than economic dynamics."[5] Marshall repeated this statement or its equivalent on numerous occasions; it offers a fair introduction to his outlook on process. The nature of "economic biology" is not treated extensively by him, but occasional statements on the subject (mainly in prefaces), plus his efforts to analyze specific social processes, give a fairly detailed picture of what he meant by the term. He offered discussion of the matter under the head of three main principles: (1) the organic analogy, (2) the principle of continuity, and (3) "the many in the one, the one in the many" (the motto of *Industry and Trade*).

I. THE ORGANIC ANALOGY

Marshall was much given to reasoning by analogy: and he wrote considerably about the relative merits of various types of analogies in the social sciences, comparing the mechanical, the chemical, the biological, and the anthropological. However, he indicated his awareness of the dangers involved in such reasoning.

"It has been well said that analogies may help one into the saddle, but are encumbrances on a long journey. It is well to know when to introduce them, it is even better to know when to stop them off. Two things may resemble one another in their initial stages but after a while they diverge: and then the comparison begins to confuse and warp the judgment."[6]

The *Principles* was a book on economic mechanics. Marshall chose such an "un-lifelike" analogy for what turned out to be his major work because he thought it would be more readily understood.

5. *Principles*, p. xiv; also "Mechanical and Biological Analogies in Economics," *Memorials*, p. 318.

6. "Mechanical and Biological Analogies in Economics," *Memorials*, p. 314.

He also argued that the analogy to static mechanics offered the most useful first approximation to the operation of the market because of the opportunities therein for reduction of complex problems to simple ones.

"Man's powers are limited: almost every one of nature's riddles is complex. He breaks it up, studies one bit at a time, and at last combines his partial solutions with a supreme effort of his whole small strength into some sort of an attempt at a solution of the whole riddle. . . .   The more the issue is narrowed, the more exactly it can be handled: but also the less closely does it correspond to real life."[7]

Further, Marshall regarded the mechanical analogy as useful in economics because of the "additive nature of economic forces," when those forces are viewed in simple, short-run context.[8]

"It is true that the forces with which economics deals have one advantage for deductive treatment in the fact that their method of combination is, as Mill observed, that of mechanics rather than of chemistry. That is to say, when we know the action of two economic forces separately — as for instance the influences which an increase in the rate of wages and a diminution in the difficulty of the work in a trade will severally exert on the supply of labour in it — we can predict fairly well their conjoint action, without waiting for specific experience of it."[9]

But Marshall felt that this addition of forces had marked limitations. He wrote that, "Mill exaggerated the extent to which this can be done; and he was thereby led to make excessive claims for the deductive methods in economics."[1] Elsewhere Marshall commented,

"There is a fairly close analogy between the earlier stages of economic reasoning and the devices of physical statics. But is there an equally serviceable analogy between the later stages of economic reasoning and the methods of physical dynamics? I think not."[2]

Marshall saw clearly the proposition made clear by modern dynamic economics — that statics is merely dynamics in which the "relative velocities of the things under study equal . . . zero." But he objected to the extension of the mechanical analogy into the study of process because, as he saw it, economic change is not only quantita-

7. *Ibid.*, p. 314. Cf. also *Principles*, p. 366.
8. Alan Gruchy, in his *Modern Economic Thought* (New York: Prentice-Hall, 1947), pp. 45ff., says that Thorstein Veblen's main criticism of Marshall was directed at the "additive" aspect of the Marshallian economics. This appears to the writer to be in error. Veblen recognized Marshall's concept of normal as differing from that of Mill and Cairnes, and hailed Marshall as one of the forerunners of evolutionary economics. Cf. Thorstein Veblen, "Preconceptions of Economic Science III," *What Veblen Taught*, ed. W. C. Mitchell (New York: The Viking Press, 1947), particularly pp. 139–50.
9. *Principles*, p. 771.
1. *Loc. cit.*
2. "Mechanical and Biological Analogies in Economics," *Memorials*, p. 314.

tive, but also qualitative. Marshall saw a fairly close relationship between the quantitative and the mechanical and also between the qualitative and the biological.[3]

"Mechanical analogies ought . . . not to be abandoned hastily on the ground that economic events react upon the conditions by which they were produced; so that future events cannot happen under exactly the same conditions as they did.

"But the catastrophes of mechanics are caused by changes in the quantity and not in the character of the forces at work: whereas in life their character changes also. 'Progress' or 'evolution,' industrial and social, is not mere increase and decrease. It is organic growth, chastened and confined and occasionally reversed by the decay of innumerable factors, each of which influences and is influenced by those around it: and every such mutual influence varies with the stages which the respective factors have already reached in their growth.

"In this vital respect all sciences of life are akin to one another, and are unlike the physical sciences. And therefore in the later stages of economics, when we are approaching nearly to the conditions of life, biological analogies are to be preferred to mechanical, other things being equal; the mechanical analogy, for instance, of a satellite which is moving around a planet, which is itself moving around another centre, is helpful for economic problems; and wherever helpful it should be used. But as the science reaches to its highest work such occasions become rarer and rarer, and the tone becomes more and more that of biological science."[4]

It has been implied that Marshall was merely giving lip service to the influence of Darwin and Spencer on English social science in writing these passages. But careful examination of his three major works brings forth numerous examples of the use of biological analogies in his analysis. Marshall clearly regarded the society and the economy as organic in fundamental nature.

"The change that has been made in the point of view of economics by the present generation is then not due to the discovery of the importance of supplementing

3. *Ibid.*, p. 312. Marshall used a number of pairs of terms (virtually as parallel pairs) that lead to a certain amount of confusion of meaning. Parallel with mechanical and biological, he sometimes used elementary and advanced, simple and complex, quantitative and qualitative, tactical and strategic, static and process. It is not altogether clear why (if that is what Marshall meant to convey) a problem could not be (for example) simple, elementary and quantitative, and still be approached as essentially a problem in economic biology rather than one in mechanics. For some purposes, it might be preferable to approach the problem of, say, the firm as being that of a social suborganism rather than as a mechanically acting cog in the economic machinery. By the same token mechanical approaches to process can get exceedingly complicated, and the "chains of reasoning" can get quite long. Probably Marshall intended that these parallels be loose ones and that simple problems at elementary stages of analysis, primarily static and involving "short chains of reasoning," should be treated with a mechanics type of analysis *for the most part*. Nevertheless, due to Marshall's lack of comment, one is inclined to feel that he has overused parallels, some of which fail to match with precision.

4. *Ibid.*, p. 317.

and guiding deduction by induction for that was well known before. It is due to the discovery that man himself is in a great measure a creature of circumstances and changes with them: and the importance of this discovery has been accentuated by the fact that the growth of knowledge and earnestness have recently made and are making deep and rapid changes in human nature."[5]

"Perhaps the earlier English economists confined their attention too much to the motives of individual action. But in fact economists, like all other students of social science, are concerned with individuals chiefly as members of the social organism."[6]

This progress of economics, Marshall argued, was the result of the influence of Goethe, Hegel and Comte, all progenitors of the organic analogy in the study of society. Also, economics had shared in the changes in the moral and historical sciences which accompanied the "great stride forwards" of biology in the nineteenth century — the principle of natural selection of Charles Darwin.[7] Marshall proceeded, like a Spencer in economics, to apply that principle to economic development.[8]

Marshall carried the principle of natural selection to quite considerable extremes, virtually to the extent of inheritance of acquired characteristics.[9] But on this subject, as with most, Marshall had reservations. The most obvious parallel to be drawn between biological and social selection, of course, was that of the effect of free competition in the selection of persons and firms engaged in economic activity, and the concomitant selection of types of economic activity. In Marshall's words, "If any firm became slack in enterprise, or weak in purpose, it passed away and made room for others, with but little disturbance to the industrial organism; just as a forest tree, which has lost its vigour, passes and leaves an opening, through which some strong young plant may shoot up towards the light."[1]

But he objected to that interpretation of the Darwinian principle which regarded these natural tendencies as inevitable and necessarily beneficial. Marshall was above all a meliorist, and he could accept no such counsel. Darwin, he argued, was often misinterpreted in this respect. The notion of "survival of the fittest"

5. "The Present Position of Economics," *Memorials*, p. 154.
6. *Principles*, p. 25.
7. "The Present Position of Economics," *Memorials*, p. 154.
8. Marshall was Darwinian to the end. In later life, discussing the possibility of future life, he expressed doubt, because he could not reason at what stage of the evolutionary scale after-life should arise — with the apes, primitive man, or later. Cf. *Memorials*, p. 64.
9. Cf. *Industry and Trade*, pp. 163–64n.
1. *Ibid.*, p. 315.

was not to be construed to mean that those persons and institutions which have survived are the only ones which should, or the *best* ones which could, survive in this world. Rather, it was to be interpreted as meaning that those institutions or living creatures would survive which were best fitted to survive in their environment, whatever that environment might be. "A race of wolves that has well organized plans for hunting in packs is likely to survive and spread; because those plans enable it to catch its prey, not because they confer a benefit on the world."[2]

Nevertheless, Marshall felt that the struggle for survival in economic life was salutary in the main. Even in the wolf pack one could not deny a certain kind of nobility. In this connection, he quoted Walter Bagehot as follows, "The rough and vulgar structure of English commerce is the secret of its life; for it contains the 'propensity to variation,' which, in the social as in the animal kingdom, is the principle of progress."[3]

Marshall was both cautious and ambivalent on this issue. He concluded that prima facie, one could make a case for the beneficial nature of modern developments in industry and business organization, but not a very strong case. Marshall, here, as so frequently elsewhere, maintained the middle course. He wished to counter the arguments of those who have a vested interest in crude laissez-faire, and at the same time to avoid what he regarded as the perils of radicalism. Hence, he countered Marxism by arguing against the inevitability of business engrossment beyond economic limits; he opposed the vested interests by questioning the absolutely beneficial aspect of such growth.[4]

Marshall embellished the organic analogy in many instances. To him, the firm itself was an organism. It exhibited a life cycle, like that of a tree or a human being, with infancy, vigorous youth, zenith, decay and senility followed by death. Similarly, he regarded nations as having such a history. He explained England's relative

2. *Ibid.*, p. 175.
3. *Ibid.*, p. 92n., quoted from Walter Bagehot, *Lombard Street.*
4. The Marshallian concept of "free enterprise" is only crudely analogous to the Darwinian view of the free play of natural forces. In the "free" economic world, the economic forces are hedged in by rules. "Free enterprise" does not allow for winning of the race. Once monopoly enters the picture, "free enterprise" has been abridged. Marshall would circumscribe economic action with rules of morality — no "sharp practices" — and with social protection of the weak and poverty-stricken. In the world of nature, presumably, the dominance of some species (however vicious) is a part of the natural process. If the weak cannot adopt defenses through mutation and selection, they cease to exist.

loss of industrial and commercial superiority in the late nineteenth century in part by reference to the high rates of growth of the economies of the United States and Germany, which he compared to the high rate of growth of a young boy.[5] England, he said, suffered from old age and creaking joints, but he credited the childlike vigor of the United States to the nature of its English parentage.[6]

Knowledge itself, Marshall indicated, followed a biological pattern of growth, with the principle of selection in operation. And he traced a parental lineage in the history of knowledge.[7]

And finally, in economic fluctuations:

".. . with every spring the leaves of a tree grow, attain full strength, and after passing their zenith decay: while the tree itself is rising year by year to its zenith, after which it also will decay. And here we find a biological analogy to oscillations in the values of commodities or of services about centres which are progressing, or perhaps themselves oscillating in longer periods."[8]

Despite these many references to biological parallels, Marshall occasionally seemed to lament the lack of precision that accompanied the complexity of the reality of life. As a trained mathematician, and an optimist with regard to the future of the scientific method, he unquestionably would have preferred that life be sufficiently simple to allow him to pursue the pattern of analysis of the *Principles* on into his analysis of process — to present in prose the results of a mathematician's logic.[9]

II. The Principle of Continuity

Stemming from the organic analogy which Marshall employed, and therefore very closely related to it, is what he called the "Principle of Continuity." Social evolution, like the biological evolution to which it was analogous, was gradual, making no discrete change. *Natura non facit saltum* expresses this attitude.

It seems strange, perhaps, that a volume which is devoted in the main to a treatment of economic matters in the language of mechanics should be headed with a motto which implies growth and development. But Marshall argued that it was only suitable that such should be the case for static assumptions could be made only if the underlying pattern of change were gradual. Those fundamental

5. *Industry and Trade*, p. 134.
6. *Ibid.*, pp. 103–4.
7. ". . . each new knowledge being the offspring of others that went before and the parent of many that follow." *Industry and Trade*, p. 206.
8. "Mechanical and Biological Analogies in Economics," *Memorials*, p. 318.
9. *Principles*, p. 780.

factors which were subject to continuous change could safely be impounded in *ceteris paribus* without doing great violence to the real world only if nature could be regarded as unlikely to take any leaps.

"Economic evolution is gradual. Its progress is sometimes arrested or reversed by political catastrophes; but its forward movements are never sudden: for even in the Western world and in Japan it is based on habit, partly conscious, partly unconscious. And though an inventor, or an organizer, or a financier of genius may seem to have modified the economic structure of a people almost at a stroke; yet that part of his influence, which has not been merely superficial and transitory, is found on inquiry to have done little more than bring to a head a broad constructive movement which had long been in preparation. Those manifestations of nature which occur most frequently, and are so orderly that they can be closely watched and narrowly studied, are the basis of economic as of most other scientific work; while those which are spasmodic, infrequent, and difficult of observation, are commonly reserved for special examination at a later stage: and the motto *Natura non facit saltum* is especially appropriate to a volume on Economic Foundations."[1]

Elsewhere he commented, "If the book has any special character of its own, that may perhaps be said to lie in the prominence which it gives to . . . applications of the Principle of Continuity."[2] Marshall certainly succeeded in giving this principle great prominence in the work. That prominence often is a source of confusion. Each analytical time period had its own equilibrium, but neither was the time period fixed and defined, nor was the equilibrium really a position of rest. The "normal" activity of each shorter time period was abnormal in terms of the longer time periods, and the equilibrium of the shorter time period was, except by accident, out of equilibrium relative to the longer period. These restless categories can be made logically compatible if the devices of dynamics can be brought into the picture, giving an infinite series of time periods, with determinate equilibria for each point in time, provided the time trends of demand and supply functions can be obtained. However, the possibility of thus extending economic mechanics is ruled out of the Marshallian analysis for the reasons already quoted from Marshall himself, i.e., that the pattern of change is too complex for mere quantitative treatment.[3]

Marshall found examples of the "Principle of Continuity" wherever he looked. Not only did he find its expression in the history of

---

1. *Principles*, p. xiii. By implication, then, political events are not governed by nature or by natural law, for if nature makes no leaps, and if politics is capable of catastrophic discontinuity, it is clear that political development is fundamentally unlike economic development, or perhaps even in conflict with it.

2. *Principles*, p. vi.

3. Logical difficulties enter as well, particularly in consideration of the possibility of increasing returns. Cf. *Principles*, Appendix H.

commerce and in industrial development, but he was also impressed by its ubiquity in logical problems, in problems of definition, and even in defining the boundaries of economics. In economic analysis, Marshall found taxonomic problems which he regarded as being very similar to those which the biologist had faced as the evolutionary concept became more prevalent in his science. The very problems of deciding what constitutes economic subject matter and what laws are economic laws among social laws generally, he discovered brought to light a disconcerting lack of discreteness.[4]

Of more significance here, however, is the path of development which Marshall described in his accounts of the history of commerce and industry. He regarded the commercial and industrial hegemony of England as merely an extension of a gradual process of adaptation of means to ends that had taken place as society progressed from a more primitive to a less primitive state. Ultimately, at the modern level, this process was culminating in the industrial revolution of the nineteenth century. That process of adaptation appears, in Marshall's writing, to be one of general applicability to all ages and to all places. Western society had accelerated that process as a result of having applied rationality in the solution of problems to a much greater extent than any other culture; but even in societies where custom has for many generations dictated methods of production and commercial and trading techniques, Marshall saw, in a sort of slow-motion process, the same kind of adaptation. This view of the pattern of progress illustrates Marshall's intellectual debt to Herbert Spencer, who described the "development of society" from the most primitive community to nineteenth century English society in terms of increasing division of labor.[5]

Marshall was confident that a custom that was fundamentally irrational could not endure. Lags in development might be long; the process of change in a society ruled by custom might even be virtually imperceptible to those who do not examine closely the history of the culture or the economy in question. But adaptation was bound ultimately to mold custom to rational means of doing things. Marshall likened this plasticity of culture to that of a glacier,[6] which must ultimately conform, in some fashion, to the topography of the land on which it lies. He seems to have reasoned that since a custom is

4. "... there is continuous gradation from social laws concerned almost exclusively with motives that can be measured by price, to social laws in which such motives have little place." *Principles*, p. 33.

5. Herbert Spencer, *Social Statics* (London: John Chapman, 1851). pp. 453–55.

6. *Industry and Trade*, p. 198.

merely a time-honored way of doing things, it could not endure unless it ultimately conformed to the hard fact of economic and social rationality, for otherwise it would cease to be thus honored.

Marshall used India as his main example of such a "slow-motion" culture, and argued that much could be learned by the student of European economic history by examining development there. An examination of the facts of Indian economic life would produce a "solvent which will explain much that is now unintelligible in mediaeval economic history."[7] One could not attribute an observed social action to mere custom, he argued, without admitting ignorance of its causes. Customary action, like any other human action, is explainable if sufficient facts are available. Ultimately, all custom must be traceable

". . . to the slow equilibration of measurable motives. . . . even in such a country as India no custom retains its hold long after the relative positions of the motives of demand and supply have so changed that the values, which would bring them into stable equilibrium are far removed from those which the custom sanctions."[8]

Marshall regarded this process of adjustment of means to ends as cumulative,[9] and as a consequence, it was much more highly developed and more readily observable in the more "advanced" economies of the West. This cumulative adaptation gradually became a more conscious type of adaptation. It was evidenced in Western society as the "business point of view."[1]

Hence, it appears that Marshall had some notion of culture growth in which rational solutions to material problems compounded into greater rationality, and consequently economic progress has moved onward and upward at an ever increasing pace. Just why those economies which are not primitive and not Western have not progressed as far as the Western economies Marshall did not explain, except to note that tropical climes are inimical to development of activities and personality traits which are conducive to a high rate of economic progress.[2] In any case, rationality was fundamental to human action in Marshall's view.

Marshall made occasional references to cyclical patterns in

7. "The Present Position of Economics," *Memorials*, p. 171.
8. *Ibid.*, p. 170.
9. *Industry and Trade*, p. 199.
1. *Industry and Trade*, p. 163. Marshall stated, "That phrase could not have been understood in a primitive society: and there is a sense in which it may be argued that business operations are merely one drift of a tendency to adapt means to ends, which is universal throughout all forms of life."
2. Marshall retained hope for the tropics in the future, however, based on the prospects of the development of solar energy. Cf. *Industry and Trade*, p. 159.

fluctuations in the history of social development. His use of the organic analogy gave him a cycle of development at one level of analysis, viz., the life cycle of the individual economic organisms (i.e., the firm, the nation, etc.). He also envisaged a pattern of cyclical movement westward in the history of civilization. "The great law of the world's history has been that energy has moved westwards, dying down in each place after a time under the influence of tyranny, and then causing new energy to spring up from its decay to the west of it: just as in the fairy ring on the grass as each ring dies down a wider ring springs up from its decay."[3] But in the main, economic progress was unidirectional for Marshall. Commercial and industrial gains were not lost, but were passed to succeeding generations. This is a rather myopic view of history, undoubtedly due to his overwhelming concern with the industrial revolution in Europe in the nineteenth century. Apparently, Marshall regarded ancient progress in commerce and industry as being of little significance beside towering modern accomplishments. His conception of progress showed a tendency on his part to interpret the past too much in terms of the nineteenth century.[4]

The industrial revolution was not regarded by Marshall as an exception to the natural continuous pattern of growth. Rather, it was regarded by him as the quite inevitable development of the broadened base of economic rationality. As the base of industrial and commercial knowledge broadened, so had the opportunity for additional advance. He often noted that as the field of knowledge grows, the area of conscious ignorance also grows.[5] This broadening of opportunity was manifest in the ever increasing scope of the economic evolution. He noted that, great as were the inventions and developments in the industry and trade of the eighteenth century,

3. "Water as an Element of National Wealth," *Memorials*, p. 140. Here is an example of Marshall's view of conflict between economic and political institutions (see n. 1, p. 585, above). This conflict suggests the technology-institutions conflict in Veblen's theory of development. In another context, Marshall spoke of a rise and fall in exceedingly long time: "My own thoughts turn more and more on the millions of worlds which may have reached a high state of morality before ours became habitable, and the other millions of worlds that may have a similar development after our sun has become cool and our world uninhabitable." *Money, Credit and Commerce*, p. 101n.

4. For example, Marshall spoke of agricultural pursuits as being inimical to rapid progress, because rigid property relations tend to develop where agriculture is dominant, and also because agricultural problems are too piecemeal and empirical for broad advancement of theoretical knowledge. Lack of economic progress in the past he credited to the overwhelming importance of agriculture in primitive society. Cf. *Industry and Trade*, p. 200.

5. E.g., *Industry and Trade*, p. 673.

those of the nineteenth were of even greater significance.[6] He clearly expected that the possibilities for coming generations would be similarly improved, provided that the great classic hurdle of diminishing returns in agriculture could be overcome.

He treated the process of invention explicitly in these terms. Although one could point to the appearance of a certain type of industrial device (or business organization) in a form that was economically usable as a landmark of development, one must understand that such developments were always social products, and never really the work of one man, or even of one age. Most important ideas could be traced for many generations in the thinking and writing of the intelligent persons of the past: and an "invention" merely marked the date of its economic fruition. The "inventor" was only the person or persons in whose hands the innovation happened to develop. Marshall noted that often such inventions occurred simultaneously in the workshops or laboratories of several men. That more inventions occurred in the eighteenth and nineteenth centuries than in previous comparable time periods was to be expected, considering the broad groundwork of ideas that was laid in the Italian, Dutch and English revivals of learning, and the concomitant expansion of industrial and commercial activity.[7]

But Marshall, although basically optimistic in his view of economic development,[8] was apprehensive lest his historical "perpetual motion machine" be hampered by the Malthusian obstacle. Marshall regarded increasing returns as the greatest single principle of modern economic development, but the marvelous gains of the industrial

6. *Principles*, p. xii.

7. Marshall regarded accumulated knowledge as being the resource of paramount importance to society. "It has been well said that if all the material wealth in the world were destroyed by an earthquake, leaving only the land, knowledge, and food enough to sustain life till the next harvest, mankind would in a generation or two be nearly as prosperous as before; but, if accumulated knowledge were destroyed, while the material wealth remained, several thousand years might be needed to recover lost ground." "Social Possibilities of Economic Chivalry," *Memorials*, p. 338.

8. "In every age of the world people have delighted in piquant stories, which tell of some local or partial retrogression; but, if we look at the broad facts of history, we find progress." "The Future of the Working Classes," *Memorials*, p. 116. This optimistic statement contrasts with the following statement by Thorstein Veblen: "But history records more frequent and more spectacular instances of the triumph of imbecile institutions over life and culture than of peoples who have by force of instinctive insight saved themselves alive out of a desperately precarious institutional situation, such, for instance, as now faces the peoples of Christendom." Thorstein Veblen, *The Instinct of Workmanship* (New York: The Viking Press, 1946), p. 25.

revolution and those that seemed to be promised by the many further opportunities in mechanics, business organization and marketing did not cause him to lose sight of the classical principle of diminishing returns. In fact, the very success of the industrial revolution served to hasten the day of reckoning with the limitations of economic space.

"Alas! There is one great hindrance. One of the first uses we are making of our increased knowledge is, as it ought to be, to save from disease and want multitudes, who even a few years ago, would have sunk under their influence."[9]

Marshall did not despair of the problem, however. He merely insisted that it was a problem which inevitably must be met and dealt with. The means he suggested were mainly those of Malthus, viz., education of the working classes and late marriage. Marshall harbored hopes that increasing returns could be realized from government investment in mass education, and exhorted his contemporaries to accept it as a principle.[1] Despite his perpetuation of this great pillar of the "dismal science" of the classical economists, however, Marshall could not be pessimistic. He merely pointed to growing population as an obstacle and exhorted men of good will and intelligence to face up to it.

III. The Many in the One, the One in the Many

This phrase was chosen by Marshall as the motto for the second volume in his series of three, *Industry and Trade*. Like the one chosen for the *Principles*, it dovetails closely with his conception of social process as a life process. Marshall conveyed in this motto his impression of the complexity of nature, of the way in which one phenomenon was causally related to numberless preceding phenomena: hence, the many in the one. Broad groups of widely differing phenomena often illustrated common characteristics which revealed the existence of a "Fundamental Idea," or common principle: hence, the one in the many.

Marshall observed that one could find, with very little biological research, examples to fit this picture. As one would trace progressively farther backward the heritage of animal or plant, the more complex would that heritage appear. Hence the more impressed one would become with the "many in the one." On the other hand, one could look at widely different types of organisms (e.g., birds and quadrupeds),[2] and despite their great differences, study would reveal certain common aspects.

9. "The Future of the Working Classes," *Memorials*, pp. 116–17.
1. *Ibid.*, pp. 117–18.
2. *Principles*, p. viii.

Marshall found an analogous set of relationships among social phenomena. In order to understand a present-day social phenomenon, one must examine its history to discover its many causes. And one must study broadly among many social phenomena to discover the fundamental ideas which link them.[3] The importance of this view in understanding Marshall's work is that it points up his belief in the need for broad empirical study to complement the abstraction process. His intention in *Industry and Trade*, as well as in *Money, Credit and Commerce*, was to make an exhaustive study of many areas of economic activity, in the hope that the "causes of causes" might be more readily understood, and, to some extent at least, controlled.

Marshall was much impressed with the accomplishments of the German and English historical schools, even suggesting that their work was one of the greatest contributions to economics in his generation.[4] He was at some pains to analyze that contribution and to offer his suggestions for integrating their work into economics. As has been frequently noted, he had a strong distaste for disputes, and he took every opportunity to pour oil on the troubled waters of the *Methodenstreit* between the historical schools and the Austrian subjectivists. It is doubtful if he added much aside from the placating influence of a great man in the field, since his pronouncements on the issues involved were frequently ambivalent and confusing. Marshall was not unaware of the problem involved in the struggle over method in economics. He pointed out that the historian could not really be as atheoretical as Schmoller sometimes suggested in the heat of controversy.[5] One could not even order historical fact, argued Marshall, without the implication of *post hoc, ergo propter hoc*.[6] Hence, the economic historian was bound to be the proponent of some kind of a theory of development, be it implicit or otherwise. The implication is that it should be explicit and carefully reasoned.

"If . . . the economic historian aims at discovering *the hidden springs of the economic order of the world*, and at obtaining light from the past for guidance in the present, he should avail himself of every resource that may help him to detect real differences that are disguised by a similarity of name or outward appearance, and real similarities that are obscured by a superficial difference. He should strive to select the true causes of each event and assign to each its proper weight; and above all to detect the remoter causes of change."[7]

3. *Industry and Trade*, Book I, chap. I.
4. *Principles*, p. 767.
5. *Ibid.*, p. 775.
6. *Loc. cit.*
7. *Ibid.*, p. 776 (Italics mine).

To elucidate, Marshall compared the problem of historical generalization to the study of naval strategy:

"Corresponding to tactics are those outward forms and accidents of economic organization which depend upon temporary or local aptitudes, customs and relations of classes; on the influence of individuals; or on the changing appliances and needs of production. While to strategy corresponds that more fundamental substance of economic organization, which depends mainly on such wants and activities, such preferences as are found in man everywhere: they are not indeed always the same in form, nor even quite the same in substance; but yet they have a sufficient element of permanence and universality to enable them to be brought in some measure under general statements, whereby the experiences of one time and one age may throw light on the difficulties of another."[8]

This distinction between strategy and tactics Marshall likened to the distinction between the uses of mechanical and biological analogies in economics.[9] The suggestion is clear that he felt the organic analogy to be the basic theoretical guide to economic history. Marshall implemented this attitude toward a theory of history for his own purposes with extreme caution, however. There are numerous examples of the application of the organic analogy to the economic growth of the business firm, the national economy and the advance of industrial technique (referred to above).[1] But Marshall did not trust deduction far in economic history. He insisted on what he called "short chains of reasoning." He would not indulge in the weaving of a broad theory of history of the Marxian type. Rather he would concentrate on the "short chains" and the "single connecting links."[2] As much as possible, Marshall wished to rely on the tried and trusted deduction of economic theory to interpret the facts of economic life. He likened certain problems in economics to some processes in industry which are simple and repeating. For that kind of problem, "mechanical" devices of analysis could be employed, and the economist could simply turn a logical crank and produce solutions. But there were problems of larger temporal scope and of larger complexity which had to be handled with method that was not mechanical. Marshall was evidently vague in his own thinking as to what that procedure was to be. For some problems in history, "common sense and mother-wit"[3] would be sufficient; but the social organism was vertebrate and therefore had to have a backbone of theory.[4]

8. *Ibid.*, p. 777.
9. *Loc. cit.*
1. Cf. pp. 582ff., above.
2. *Principles*, p. 773.
3. *Ibid.*, p. 778.
4. *Ibid.*, p. 769.

Here is a point at which Marshall's compromising outlook leads to veritable logical breakdown. He argued that one could not avoid implication of a theory of history, but he really felt that no general theory of history could be written. He never offered an explicit one of his own nor ever really analyzed the implicit theoretical factors in his personal attitude toward history. In discussing the matter, he ranged from a "great law of the world's history"[5] of a fairly explicit, teleological nature, to what might be described as a "seamless web" view. The social organism was occasionally supplied by him with a suggestion of vertebrae, but on most occasions in his hands it resembled most nearly a jellyfish. Evidence of the latter is to be found in his assertion to the effect that history is useful only insofar as the records are "sufficiently ample to suggest their own explanation."[6] And also, " . . . only a careless explanation can overlook any influence by which the result has been actually fashioned."[7] This suggestion would seem less than adequate if one's ambition is to discover "the hidden springs of the economic order of the world."[8]

J. M. Keynes put it somewhat more gently, saying that Marshall was "of two minds" about this matter:

"On the one hand his view about the perpetually changing character of the subject-matter of Economics led him to attach great importance to the historical background as a corrective to the idea that the axioms of today are permanent. He was also dissatisfied with the learned but half-muddled work of the German historical school. On the other hand, he was afraid of spending too much time on these matters (at one period he had embarked on historical inquiries on a scale which, he said, would have occupied six volumes), and of overloading with them the essential matter of his book. . . . [Everything in the historical sections of the *Principles*] is boiled down into wide generalisations, the evidence for which he has not space to display."[9]

In practice Marshall seems to have regarded history as mainly a tool for maintaining perspective with reference to current economic problems. One was to look into historical fact in order to arrive at some notion of the pattern of growth of an institution which one was interested in for some reason. "It is necessary to look backwards a

5. "Water as an Element of National Wealth," *Memorials*, p. 140.
6. *Industry and Trade*, p. 7.
7. *Loc. cit.*
8. *Principles*, p. 776.
9. *Memorials*, p. 46. Keynes also quoted Mrs. Marshall to the effect that Marshall ran into similar difficulties with *Industry and Trade*, starting an historical treatment which was on such a large scale that he had to abandon it as too ambitious. He subsequently used fragments of this "White Elephant" (as he called it), however. *Ibid.*, p. 52.

little in order the better to look forwards," he said.[1] He exercised his "common sense and mother-wit" rather than any theory of development to find historical fact to illustrate those developments of Western history which his predilections indicated to be of overriding importance, viz., the industrial revolution of the nineteenth century and the rise and relative decline of English hegemony in industry and commerce.

In Marshall's defense, however, this much should be said. There is a significant difference between being explicitly aware of a problem which one fails to meet satisfactorily (as appears to be the case with Marshall on economic development), and lacking awareness of the problem itself. Marshall was bold enough to recognize and assert that the economist's focus should be on dynamic reality rather than on static logical reasoning, but he saw no way of handling the problems of change with care and rigor comparable to his static model. Nevertheless, he devoted countless hours to study and writing on historical matters, and it appears that those who have interpreted him are responsible for giving the erroneous impression that he saw the economic world as a static mechanism.

Although Marshall did not come forth with a detailed theory of history, he exhibited in his attitude toward economic development some of the symptoms of "historicism"[2] which are usually attributed to those writers of history who employ some form of the organic analogy. There are fairly numerous indications of a teleological or deterministic outlook. One such indication, already referred to,[3] is his faith in progress. "If we look at the broad facts of history," he said, "we find progress."[4] Also, Marshall made a vague reference to a stage process in the development of science. " . . . if the subject matter of a science passes through different stages of development, the laws which apply to one stage will seldom apply without modification to others: the laws of the science must have a development corresponding to that of things which they treat."[5] The context seems to imply a reference to Comte's stages in the development of

1. *Industry and Trade*, p. 5.
2. "Historicism" is the term applied by Karl Popper in his attack on the extremes among theories of history. In general, the epithet applies to the aspect of teleological inevitability to be found in such theories. Cf. K. R. Popper, *The Open Society and its Enemies* (London: Routledge and Kegan Paul Ltd., 1949).
3. See p. 589, above.
4. "The Future of the Working Classes," *Memorials*, p. 116.
5. "The Present Position of Economics," *Memorials*, p. 154. Marshall indicated a predilection for stages in a number of contexts. Most explicit is his acceptance of the stages of national development suggested by Friedrich List. Cf. *Industry and Trade*, pp. 698–99.

Positive science. In another passage, Marshall's view of the progress of science reflects quite clearly the determinism characteristic of Spencerian and Comtist thought: "The change [in social science] is mainly due to the irresistible forces of the age affecting at once all the rising generation in all parts of the world."[6]

Marshall's preoccupation with economic matters inevitably led him to virtual economic determinism in his interpretation of the progress of humanity in the sense that economic problems were deemed the basic human problems. The process of adapting means to ends throughout history, Marshall suggested, has been primarily one of economizing. This process of adaptation has shaped the development of custom and social organization as well as of industrial and commercial technique.

Whether Marshall's outlook can be called deterministic or not, it is overwhelmingly optimistic. His life work was devoted to the furthering of the process which he felt was at work in his time, i.e., the raising of the plane of living of the average man above the poverty level, so that he might become a social asset rather than a social liability. In this process, he was sure that further progress could occur, if man would make a serious, conscious effort to bring it about

Bruce Glassburner.

University of California
Davis, California

6. *Memorials*, p. 154.

THE EXCHANGE BETWEEN QUANTITY AND QUALITY*

By Jack Hirshleifer

I. Introduction

In economic analysis, the cost isoquant[1] between two different commodities from the point of view of a purchaser or a producer is in the general (possibly noncompetitive) case assumed to be concave to the origin — the limiting case being that of a purchasing unit with no monopsonistic power forced to move along a budget curve which is a straight line. In a recent paper[2] H. Theil has pointed out that under certain conditions the choice between *quality* and *quantity* of a commodity[3] involves a consumer's budget line which is *convex* to the origin. This is illustrated in Figure I, in which n refers to numbers or quantity of the commodity at any quality level, q is a measure of quality, C is the budget or cost isoquant, and U a consumer's indifference curve. Mr. Theil also points out that this convexity leads to instability of the tangency solution if it should happen that the utility isoquant is of lesser curvature than the cost isoquant; when this is the case, the point of tangency will represent the worst rather than the best position along any given isoquant. In a problem

* The author would like to express his indebtedness to Harry Markowitz for his valuable criticisms and suggestions. Remaining errors are, however, solely the responsibility of the author.

1. This term will be used generally to refer to any or all of the iso-cost curves which arise in different branches of economic theory, such as the transformation or production possibility curve in the theory of the firm or of international trade, and the budget line in consumer theory. There is, unfortunately, no single term covering the utility function for the consumer and the revenue function for the firm. In the absence of such an expression, we shall use "utility function" throughout, meaning utility for the consumer or revenue for the firm, as appropriate.

2. H. Theil, "Qualities, Prices, and Budget Enquiries," *Review of Economic Studies*, XIX (3), No. 50 (1952–53), 129–14.

3. To avoid awkward circumlocution we shall use the term "commodity" throughout to refer to a group of similar products, rather than a unique good. A Cadillac and a Chevrolet are then different qualities of the commodity "automobile."

in which the author has been engaged, a comparable situation led to a convex cost isoquant between quality and quantity of product in the theory of the firm. It appears that quite plausible assumptions may lead to convex cost isoquants for the consumer, firm, or economy
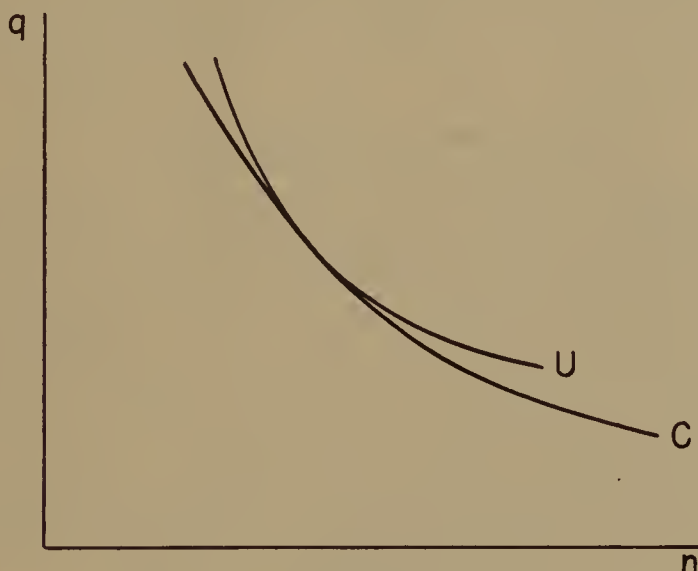

FIGURE I

in spite of the fact that the marginal rate of substitution behaves rather strangely for such isoquants. The main purposes of this paper are to look into the question of how general such isoquants and resulting instabilities may be in the quality-quantity problem, considering not merely the possible mathematical formulations of the cost function but also taking into account economic intuition.

The question of quantity versus quality, though slighted by traditional theory,[4] is in fact one of the most universal problems in the practical application of economic theory. A manufacturing firm may, at a given total cost, be able to produce a small number of better machines or a larger number of a more primitive model; a consumer may be in a position to choose between more food or better food; or a military department of government may be faced with a choice between procuring small numbers of individually excellent

4. A rather considerable literature has by now developed, however, as an outgrowth of monopolistic competition theory, upon various aspects of the broader question of product variation. A useful bibliography will be found in E. H. Chamberlin, "The Product as an Economic Variable," this *Journal*, LXVII (Feb. 1953).

weapons or larger numbers of more commonplace ones. For each of these cases there will be an iso-cost or budget line connecting points of equal cost; the "utility" isoquant of Figure I will be in the first case a revenue isoquant, in the second a utility indifference curve, and in the third an isoquant connecting points of equal military value.

It is convenient to use isoquant analysis for the quality-quantity problem provided it can be assumed that, while there is a possible choice between different models of the "same" commodity, only one model will actually be procured — that is, mixed purchases are not considered. While this assumption restricts the applicability of the analysis, such an assumption can often be safely made in certain practical problems.[5] Ordinary isoquant analysis, in which differing quality levels are treated as different commodities, is more general in that mixtures are taken into account without difficulty. On the other hand, discrete levels of quality rather than a quality continuum must be assumed. A continuum may be approximated by considering a sufficiently large number of discrete cases, but in practice such an analysis will become awkward when more than three, or even merely three, quality levels have to be considered. Therefore, while in principle the problem can be handled as a special case by ordinary commodity analysis, the quality-quantity technique outlined here will often be the only practicable method of solution.

It is desirable at this point to clarify somewhat the use of the term "quality," not heretofore defined in this paper. We do not want to reduce the idea of quality to mere subjective preference, since the purpose of this note is to show that the intuitive concept of objective quality can be quite useful in applications of economic theory. What seems to lie behind the intuitive concept is the idea that commodities may be grouped into families according to the *purposes* which these commodities ordinarily serve: some commodities provide transportation, others heat, others food, and so forth. Within a large category like transportation, we have subfamilies of commodities providing air transport, land transport, sea transport, and so forth. When the subfamily is sufficiently narrowly defined, it frequently becomes possible to find some objective technical characteristic which is, at least in the *ceteris paribus* sense, positively associated with performance — i.e., achievement of the purpose ordinarily served by the class of commodities in question. For the subfamily of commodities "transport aircraft," for example, reasonable

5. Alternatively, one can take mixtures into the analysis by using the concept of average quality of a purchase, provided it can be assumed that the average sufficiently describes the mixture for all purposes.

though partial quality characteristics are cargo capacity, air speed, durability, etc. In almost all practical problems, quality would have to be represented as a vector of such characteristics, though it frequently happens that the conditions of the problem preclude variation in more than one or only a few of the conceivably variable characteristics. In this paper, we shall assume that only one quality characteristic can be varied, so that it alone serves as a sufficient measure of quality for our problem.

II. The Convex Cost Isoquant

Our purpose in this section is to explore the conditions leading to convex cost isoquants in the quality-quantity problem. Let us note here what a convex cost isoquant implies. It means essentially that movement at a given cost toward either extreme — very high quality with small numbers, or vice versa — can be made on more and more advantageous terms as you approach the extreme. In ordinary economic situations, as in the production of two commodities, we generally assume that the marginal rate of substitution increases (shifts unfavorably) as you push toward extreme combinations. More specifically, a convex cost isoquant means that instead of it becoming harder and harder to get additional quality (in terms of what you must sacrifice in quantity in order to keep cost constant) as you move from combinations already involving high quality and small numbers (and vice versa for the opposite case), it becomes easier and easier — until in the limit one can buy an indefinitely large increment of quality with an infinitesimal sacrifice of quantity. Economic intuition does perhaps rebel against convex cost isoquants, and yet the conditions of the following examples may seem reasonable enough.

Let C = total cost, C_0 = fixed cost,[6] and c = average variable costs. As before, n refers to numbers or quantity, and q to quality.

$$(1) \quad C = C_0 + nc\,(q,n)$$

$$\frac{\partial C}{\partial n} = n\frac{\partial c}{\partial n} + c$$

$$\frac{\partial C}{\partial q} = n\frac{\partial c}{\partial q}$$

6. Where the commodity in question is purchased rather than produced, as in the case of the consumer, there will be no fixed cost term.

Along any cost isoquant,

$$\frac{dq}{dn} = -\frac{\dfrac{\partial C}{\partial n}}{\dfrac{\partial C}{\partial q}} \cdot$$

It is reasonable to assume that:

$$\frac{\partial c}{\partial n} \geqslant 0, \; \frac{\partial c}{\partial q} > 0, \; \frac{\partial^2 c}{\partial n^2} \geqslant 0, \; \frac{\partial^2 c}{\partial q^2} \geqslant 0$$

Routine steps will show that the sign of $\dfrac{d^2q}{dn^2}$ will be that of the following expression, of which the first term is positive, the next two indeterminate, and the remaining four nonpositive:

$$2\,c\,\frac{\partial c}{\partial q} + 2\,cn\,\frac{\partial^2 c}{\partial q\,\partial n} + 2\,n^2\,\frac{\partial c}{\partial n}\,\frac{\partial^2 c}{\partial q\,\partial n} - n^2\,\frac{\partial c}{\partial q}\,\frac{\partial^2 c}{\partial n^2} - 2\,cn\,\frac{\dfrac{\partial c}{\partial n}\cdot\dfrac{\partial^2 c}{\partial c^2}}{\dfrac{\partial c}{\partial q}}$$

$$- c^2\,\frac{\dfrac{\partial^2 c}{\partial q^2}}{\dfrac{\partial c}{\partial q}} - n^2\,\frac{\left(\dfrac{\partial c}{\partial n}\right)^2 \dfrac{\partial^2 c}{\partial q^2}}{\dfrac{\partial c}{\partial q}} = \frac{d^2q}{dn^2}\,n^2\left(\frac{\partial c}{\partial q}\right)^2$$

It is evident that we cannot make any simple general statement about the sign of the second derivative. However, the reader may verify that such a plausible average cost function as

$$(2) \quad c = a\,e^{bqn}, \qquad (a,\, b > 0)$$

will lead to a positive $\dfrac{d^2q}{dn^2}$, and so to a convex cost isoquant and diminishing marginal rate of substitution, even though there are increasing average and marginal costs with respect to both quality and quantity.[7] For the limited special class of cases where constant

---

7. For example, let q = maximum altitude capability in feet for fighter aircraft, and n = number of fighter aircraft produced in a specified time period. The function (2) shows increasing marginal and average cost, for any specified altitude capability, as numbers increase, as well as increasing costs for any given quantity as altitude capability increases. The quantity-altitude cost isoquant will, however, be convex.

costs to quantity may be assumed, it is possible to determine rather broad conditions on the cost function which lead to convexity.[8]

Convexity of the cost isoquant may hold, therefore, under a certain not unreasonable class of conditions in the quality-quantity problem. It has occurred in practical applications in the author's experience where equation (2) has been supplied by an engineer or production specialist. It should be recalled that, since we have been ruling out mixtures by assumption, this convexity has no relation to the possible advantages of specialized as opposed to mixed production.

Our emphasis on convexity of the cost isoquant in the exchange between quality and quantity should not be taken to imply that convexity is impossible in the exchange between quantities of two commodities. Convexity will occur in this exchange, under the usual assumptions of positive and increasing marginal cost, if the cross-derivative of the total cost (i.e., the change in the marginal cost of commodity x as the quantity of commodity y is increased, or the change in the marginal cost of commodity y as commodity x increases) is sufficiently positive.[9]

It is customary to assume, however, that multiple products are complementary in production, which implies a negative cross-derivative and thus assures a concave cost isoquant.[1] Or, if the commodities are purchased rather than produced, it is most reasonable to assume a zero cross-derivative — and again, convexity of the cost isoquant is ruled out. In contrast, the different form of the relationship between the elements of cost in the quality-quantity problem permits convexity to hold under quite plausible conditions.

8. For example, if average variable costs are a function only of q, $\dfrac{d^2q}{dn^2}$ is positive, zero or negative as $\dfrac{d^2}{dq^2}\left(\dfrac{1}{c}\right)$ is positive, zero, or negative. That is, the cost isoquant will be convex or concave as the reciprocal of $c(q)$ is convex or concave. It can be shown that convexity will hold, under these assumptions, for many simple types of mathematical functions (e.g., polynomials and exponentials), at least for q sufficiently large. The proof will not be given here.

9. If $C = \phi(x,y)$, and assuming $\phi_x, \phi_y, \phi_{xx}, \phi_{yy} > 0$,

$$\frac{dy}{dx} = -\frac{\phi_x}{\phi_y},$$

and a convex cost isoquant occurs if and only if

$$2\phi_x\phi_y\phi_{xy} > \phi_{xx}\phi_y^2 + \phi_{yy}\phi_x^2$$

For an analogous discussion of the utility isoquant, see G. J. Stigler, *The Theory of Price* (Rev. ed., 1952), p. 301.

1. See J. R. Hicks, *Value and Capital* (2d ed., 1946), pp. 97–98. But cf. Stigler, *op. cit.*, pp. 129–32.

III. Equilibrium Conditions

As has already been noted, the point of tangency of the cost and utility isoquants will be a nonoptimal solution if the former isoquant is more convex than the latter. From an a priori point of view, there seems to be no reason to believe that any particular convex curvature is more likely than any other, and so the practical possibility of failure of the tangency solution cannot be rejected.[2] This situation may lead to a boundary solution, but in a wide class of cases we shall see that such a relation between the isoquants cannot hold throughout the graph and so a stable tangency solution exists.
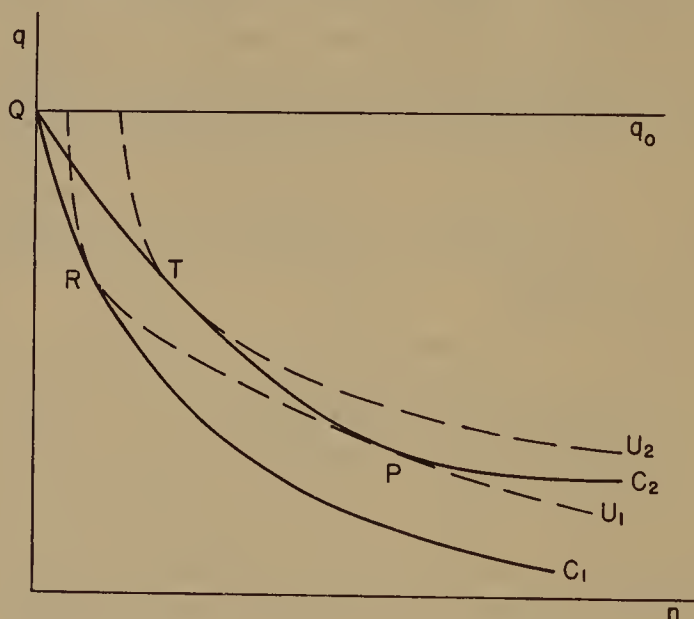
Figure II illustrates a case in which there are convex cost isoquants, C_1 and C_2, a limiting value q_0 on quality, and a nonoptimal tangency solution with the utility isoquant U_1 at P, where the curvature of U_1 is less than the curvature of C_2. The crucial point turns out to be the behavior of the cost function as q approaches its limit q_0. To clarify this point it will be convenient to cite examples. If the ultimate level of quality can be approached but not actually attained, unit cost would be expected to go to infinity at $q = q_0$. This might be the case if our measure of quality were speed of a material object — either proceeding under its own power or accelerated by a gun or a

---

2. We continue to assume that the utility isoquant is of the normal convex curvature.

cyclotron — since, at least under current physical theories, the limiting velocity of light can only be approached. (This limit may not be of practical importance for problems like surface transportation, but for particles in a cyclotron it definitely is significant.) In the other case, the ultimate level of quality may be quite attainable, possibly at high but not at infinite cost. This might occur in practical problems if the quality characteristic in which we were interested was the absence of living bacteria in, say, a sample of milk or water — since, for reasonably sized samples, techniques are available which will kill literally all bacteria in the sample.

In the former case, which is the one illustrated in Figure II, all the cost isoquants (for $C_i > C_0$) will converge on the point Q where $n = 0$, $q = q_0$, though none of them will actually be defined at Q. In the latter case, the cost isoquants will have quite a "normal" appearance along the line $q = q_0$ (see Figure III). In this case, it
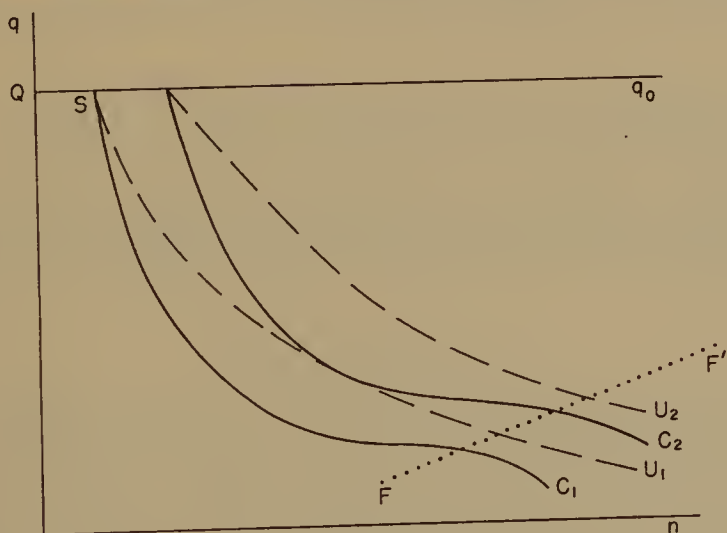
FIGURE III

is quite clear that, as illustrated in Figure III, there may be no stable tangency solution at all; with sufficiently convex cost isoquants, all the solutions may, like the point S, lie along the boundary $q = q_0$. For the case illustrated in Figure II, on the other hand, it is evident from the geometry that there will be stable tangency solutions at R and T as well as the unstable solution at P, provided that the utility isoquants intersect the line $q = q_0$ to the right of Q.

The definite tendency toward a boundary solution at the ultimate limit of quality where the cost isoquants are of the form of Figure III — that is, where the ultimate level of quality is attainable

at finite cost but the cost isoquants are more convex than the utility isoquants — is perhaps of practical importance, though the author is not aware of any particular class of phenomena showing this pattern.

The cost isoquants of Figure III have also been purposely drawn to illustrate another possibility, one which has actually occurred in a practical problem. In this case the cost isoquants are convex in the high-q low-n region but change to the "normal" concave shape in the high-n low-q region. For the curves as drawn, the line connecting optimal solutions runs along $q = q_0$. A small change in one or more of the parameters, however, will suffice to make solutions along the line FF' dominate these. In practical language, a comparatively small change making quality more expensive than before relative to quantity may produce not a marginal change in the solution, but a radical shift from a choice involving a small number of very high-quality units to one using a large number of comparatively inferior units.

Our analysis can be modified, without essential change, to cover the range of cases which occur when the measure of quality is not intrinsically bounded but can rise without limit: e.g., expected length of life of some mechanism. The analysis is complicated by the need for dealing with points at infinity, but the same classes of cases occur in principle.

IV. Utility Function

The nature of the utility function in the quantity-quality problem will only be briefly examined here. On the basis of observation of the facts of economic life, we believe it safe to make the following assumptions: (1) all combinations involving zero quantity will be on the zero utility isoquant,[3] and (2) all nonzero utility isoquants have a negative slope, with higher isoquants lying to the right and above lower ones. In addition, we believe the assumption that, (3) utility isoquants are convex to the origin, at least in the large, will be generally granted.

The most interesting assumption is related to the question of continuity of the preference relation.[4] We assume: (4) For any

3. We shall avoid having to give a definition to the concept of zero *quality* by limiting our attention to the behavior of the cost and utility functions away from the lower limit on quality, if any.

4. See J. Marschak, "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*, XVIII, No. 2 (April 1950), 111–41 (Postulate II, p. 117).

combinations A, B, C such that A is preferred to B and B is preferred to C, there is some "average" of A and C which is indifferent to B. By "average" we mean, geometrically, some point on the straight line between A and C other than the points A and C themselves. In Figure IV, we have M preferred to N (this follows from
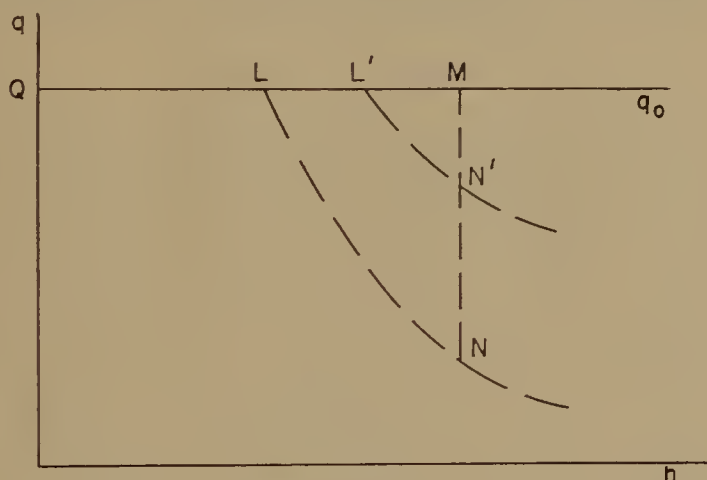

FIGURE IV

assumption 2) and N preferred to Q (assumption 1), whence it follows that there is some L between Q and M indifferent to N. Similarly, for any L' between L and M, an N' between M and N indifferent to L' can be found.

Assumption 4, or more particularly our application of it, says essentially that one traverses all levels of utility by proceeding from Q along $q = q_0$, and also all intermediate levels of utility in proceeding from M to N. That is to say, there is nothing so overpowering about the quality level q_0 as to rule out consumer indifference between a certain n along q_0 and some larger n at a lower quality level. Thus, absolutely bacteria-free milk is better than an equal volume of almost bacteria-free milk, but there is some quantity of the latter which will be indifferent to an eyedropper-full of the former.

Assumption 4 is plausible, and undoubtedly covers a wide class of cases (our utility isoquants in previous examples have all been based on it), but one might well doubt whether it holds universally. For example, to achieve a certain effect of overwhelming importance, it may be necessary to have a certain volume of milk absolutely free of bacteria — almost bacteria-free milk being usable only for secondary purposes of sharply reduced utility. This implies a discontinuity in the preference relation. An example of an isoquant map which

meets our other conditions but denies assumption 4 is illustrated in Figure V, in which all the curves of negative slope are utility iso-
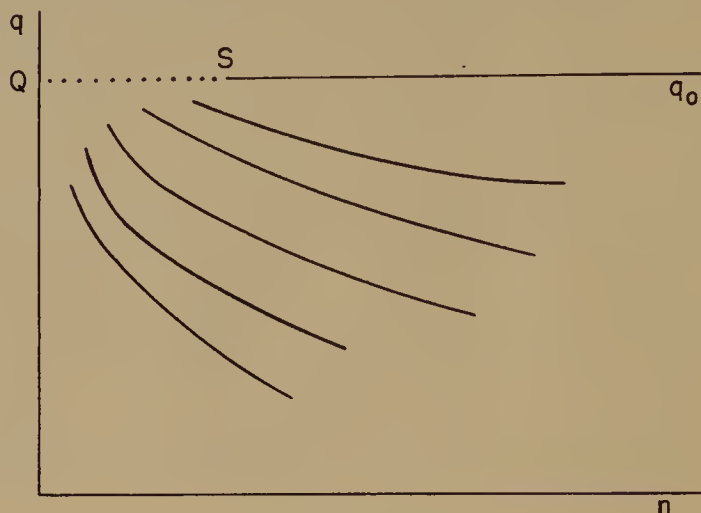

FIGURE V

quants. $\overline{QS}$ is the minimum volume of pure milk necessary to achieve the desired objective. Then all the utility isoquants drawn intersect $q = q_0$ to the left of S. The part of this line to the right of S is not a utility isoquant, but represents points of increasing utility any of which is preferable to any point in the graph below $q = q_0$.

V. Conclusion

The practical significance of the considerations raised here is that, in the exchange between quality and quantity from the point of view of the consumer or of the firm, various possibilities arise which should make the economist careful in using his intuition trained in the exchange between quantities of different commodities. We have emphasized most the possibilities of decreasing marginal rate of substitution (convexity) in the cost isoquant, which may arise under quite plausible conditions. An important possible consequence is a certain tendency for extreme solutions to occur, as in the case examined where small changes in one of the underlying parameters could lead not to a marginal modification of the solution but to a widely different outcome.

JACK HIRSHLEIFER.

THE RAND CORPORATION

SANTA MONICA, CALIFORNIA

MONOPOLY AND COMPETITION IN EUROPE AND AMERICA

By Tibor Scitovsky

Monopoly and competition have been a much discussed topic among economists ever since the publication in 1933 of Professor Chamberlin's and Mrs. Robinson's well-known works on the subject. They provided the analytical tools and theoretical framework of the subject; and it is since then that most of us think of the two concepts as merely limiting cases and constituent elements of a more general type of market behavior. In the intervening two decades most of the ideas contained in these works have been further developed, valuable new ideas have been added, and much empirical research has been done; nevertheless, our understanding of the subject has not advanced very much. This, to some extent, is a tribute to the early work on the subject; but it also reflects unfavorably on the work done since then. For one thing, the literature of the subject is very diffuse, replete with competing approaches, terminologies and classifications; many of them good and useful but creating, all together, a bewildering effect of redundancy and unnecessary variety, and rendering an integration of the various results almost impossible. For another, much of the recent work has tended to stay in accustomed grooves and accept traditional value judgments, without bothering to re-examine their validity. This is not to imply that the originators of the new approach made wrong value judgments and cut grooves leading in the wrong direction; but the last twenty years have brought a shift in the economist's interests and emphasis, to which, I suspect, few of the contemporary writers in this field have paid enough attention. It would be desirable, therefore, to review and reappraise our entire approach to the problems of monopoly and competition, and to try and develop whatever modifications or additions to accepted theory seem to be called for in view of past experience and of the change in the economist's interests and emphasis. It is gratifying to find that a first preliminary step in this direction has recently been taken.

In 1951 the International Economic Association held a round-table conference on "Monopoly, Competition and their Regulation" at Talloires, France; and the papers presented, complete with a sum-

mary of their discussion, have now been published.[1] Books of this kind are seldom very successful; but this is a welcome exception. Indeed, it is one of the few multi-authored books worth buying, even on budgets constrained by academic salaries. Much credit is due to the chairman of the Program Committee, Professor Chamberlin, and to the members of his committee, for an excellent selection of authors and topics, and a happy matching of authors with topics.

This is not to say, however, that the papers are uniform in quality. For one thing, the eight papers that describe the state of competition in eight different countries are as varied as is the relevant statistical information available in these countries. For another, of the analytical papers (which form two-thirds of the remaining sixteen) a few make important original contributions, others do little more than restate or summarize positions already known. But the value of the book lies not so much in the new ideas it contains as in the survey of the entire field it affords. It is very useful to have within the same volume, a statement of the issues involved, accounts of the empirical information available, appraisals of various attempts to deal with the monopoly problem, restatements of most of the important approaches and ideas in the field, and a "Summary Record of the Debate" to which all these were subjected during the conference. Much can be learned from the juxtaposition of competing theories, from the discussion of practical problems, and last but not least, from the failure of the conferees to resolve some of the problems discussed. The book might have been more constructive had the papers been written after and not before the conference; as it is, it provides the discerning reader both with a good survey and criticism of the accepted body of theory and with many of the elements that would have to go into an amendment of that theory.

It is not surprising perhaps that several of the papers condemn competition, both as a form of market organization and as a theoretical standard by which to judge the performance of an actual economy. Mrs. Robinson leads the attack by arguing "The Impossibility of Competition," mainly on the basis of the cobweb theorem, which she applies to the problem of investment. She is seconded by Mr. Rothschild, who presents an imposing if mostly familiar catalog of "The Wastes of Competition" under both ideal and monopolistic conditions; and these theoretical strictures on competition are well

---

1. *Monopoly and Competition and Their Regulation.* Papers and Proceedings of a Conference held by the International Economic Association, ed. E. H. Chamberlin (London: Macmillan Co. Ltd.; New York: St. Martin's Press; 1954).

supported by Mme. Aubert-Krier's empirical study of "Monopolistic and Imperfect Competition in Retail Trade" in France.

Criticism, however, is meted out impartially; and the attack on monopoly and other forms of market restriction is just as heavy if not quite as effective. Professor Schneider attacks cartels for their failure to promote research and product development, and on the grounds that they aim to maximize profits, not efficiency, and that, to judge by their frequent failure to raise the profits of the participants, they seldom achieve efficiency even as a by-product. He has a bad word to say also for monopoly: "the widely held assumption that the ownership and control of plural production by single corporate enterprises contributes to efficiency would seem to rest upon an overwhelming absence of support" (p. 210). Professor Lombardini, in his examination of "Monopoly and Rigidities in the Economic System" adds an interesting new kink (the uncertainty kink) to the familiar oligopoly kink, to explain the rigidity of monopoly price; and Professor Machlup condemns this rigidity — even when disguised as stability — by arguing that "artificial stability of particular prices aggravates the instability of general incomes" (p. 394), and that even general wage stability, due to "the exercise of monopoly power by trade unions may be extremely unstabilizing at times" (p. 393). Professor Hennipman, in an excellent paper on "Monopoly: Impediment or Stimulus to Economic Progress?", warns against too ready an acceptance of the view ascribed to Schumpeter that "the tendency of monopoly in general is conducive to progress" (p. 451); and Professor Lewis, discussing the "Recent British Experience in Nationalization," is sceptical about the efficiency of nationalized enterprises, and cautions against pinning high hopes to them as an alternative to private monopoly. Professor Jeanneney draws somewhat less gloomy conclusions from the experience of "Nationalization in France."

Needless to say, all these views were not held by each of the persons mentioned; but there seems to have prevailed among the European members of the conference a general feeling of pessimism, concerning the performance of competitive industry, the performance of monopoly, both private and public, and the possibility of controlling monopoly. Hence Mrs. Robinson's prayer for a "third way between a choice of two evils" (p. 513), which must have been shared by many other European members of the conference. The only hopeful note came from Sweden, as though in answer to Mrs. Robinson's prayer, in Professor Svennilson's discussion of the successful Swedish experience with "market therapeutics" — *ad hoc*

measures aimed at improving the organization and efficiency of particular industries, based on the recommendations of study groups appointed by the government or formed by the industry itself.

In striking contrast to the Europeans' pessimism was the confidence of the Americans concerning the competitiveness and efficiency of their own economy. Professor Heflebower, reporting on "Monopoly and Competition in the United States of America," maintained that competition in the United States, while wanting by theoretical standards, was nevertheless workable and adequate in some deeper sense of the word. Similarly, Professor Chamberlin felt that "monopolistic competition may, in practice, contain more and stronger elements of competition than the unrealistic model of pure competition" (p. 508); and Professor Clark, in "Competition and the Objectives of Government Policy," argued that if the U.S. economy was not more competitive, this was due not to our inability to check monopoly but to the American public's not wanting more competition. The Canadians seemed to share much of this optimism, at least to judge by Mr. McGregor's report on "Preventing Monopoly — Canadian Techniques."

The difference in spirit between the European and American members, and the underlying difference in performance between European and American economies, was much debated and commented upon during the conference. There were arguments that America is "full of pushful immigrants" (p. 534), stress was laid on the difference in man's aspirations between Europe's old and America's new civilization (pp. 57–58), pleas were made for research into American institutions that might explain the difference between American and European competition, American and European monopoly (p. 495); but there is no evidence in Mrs. Henderson's excellent "Summary Record of the Debate" of any explanation of the difference with the aid and within the framework of economic theory. Indeed, she says explicitly that the problem remained unresolved (p. 495); and this, I feel, is a serious indictment of the traditional theories of monopoly and competition. For it seems obvious that the superior performance of the American economy must have something to do with its having these "more and stronger elements of competition," whose existence most American members of the conference asserted, but whose nature and significance none explained. And yet, most elements of an explanation seemed to be there, like the pieces of a jig-saw puzzle, waiting to be assembled and to have the missing pieces supplied. The problem is challenging and impor-

tant enough to be pursued a little further, if only as an illustration of the thoughts aroused by, if not always contained in, the book under review.

The difference between American and European efficiency is a difference primarily in technological efficiency, which has to do with the entrepreneur's choice among alternative methods of production, with his success or failure in keeping up with scientific progress, and with specialization among different members of an industry. Efficiency in this sense is coming more and more to the forefront of the economist's interests, owing to its relevance to many contemporary economic problems, among them the problem of economic growth. Let us remember, however, that this shift in interest and emphasis is a recent development, to which many branches of economic theory have not yet had a chance to adapt themselves.

At the time when the theories of monopolistic and of imperfect competition were developed, the economic effects of competition that received the most attention were income distribution, resource allocation among firms and industries, the utilization of existing capacity, and the managerial efficiency with which the entrepreneur runs his business. All these effects hinge on (1) the relation of price to the marginal cost of products or the marginal value of factors, and (2) the tendency (through the entry of newcomers) towards the competitive elimination of monopoly profit. These two categories stand in the center of much of the recent work on monopolistic and imperfect competition, which explains its success in dealing with the problems of competition and monopoly as they relate to the above-mentioned effects.

As Professor Svennilson points out in "Monopoly, Efficiency, and the Structure of Industry," today most of the above effects have ceased to be regarded as very important. Interest in the effect of market structure on income distribution has waned, because in a " 'world of monopolies' . . . different monopolies partly neutralize each other's efforts to exploit the rest of society" (p. 272), and also because taxation and social security policy have become effective tools for mitigating inequities of income distribution. As to the other effects mentioned above, I still think they are important, but they have certainly been overshadowed by the problem of technological efficiency as defined above. Efficiency in this sense may be measured by the relation of the firm's actual cost to the socially necessary minimum cost of production, as this is determined by the technical knowledge of the time and the factor availabilities of the place; and

this relation, while known in the theoretical literature in the shape of Chamberlin's long-run cost curve, has not yet become an operating concept in the empirical work.[2] Moreover, the technological efficiency of the firm, the method of production it adopts, its ability or failure to keep up with technological progress, all have very much to do with the firm's scale of investment and operations, whence it follows that any theory trying to explain the influence of market structure on the firm's performance in the above senses must also concern itself with the relation between market structure and the scale of investment and output. The traditional theories, however, are ill-equipped for this task, primarily because their main concern is with current production, not with investment, and also because they tend, on the whole, to abstract from (long-run) problems of scale.[3] Their primary purpose, after all, was partly to explore the implications for static equilibrium theory of the relaxation of one of its simplifying assumptions, pure competition; and partly to enable equilibrium theory to deal with the problems raised by product differentiation and nonprice competition. This purpose, I believe, was pretty well accomplished. What remains yet to be done is to explore the consequences of the relaxation of yet another simplifying assumption, perfect divisibility and irrelevance of scale. This might best be done in the framework of a theory of investment under monopoly and competition, which could then be regarded as a new chapter added to accepted theory.

To explain the difference between American and European efficiency one must explain why American and European entrepreneurs pursued different investment policies and invested in different types of plant and equipment. When one realizes that, as a rule, the more efficient a method of production, the larger the scale on which it must be operated to make it profitable, one is tempted to explain everything by the larger scale of the American economy. This, however, would be a very incomplete explanation. To begin with, Professor Robinson pointed out that in most industries the size of the average manufacturing plant, when measured in terms of man power, is not significantly smaller in Europe than it is in America (p. 503). It is

---

2. See, however, the reference to Bain on p. 616 below.
3. The only exception I know of is E. H. Chamberlin's "Proportionality, Divisibility and Economies of Scale" (this *Journal*, Feb. 1948), reprinted as Appendix B in the 6th edition of his *Theory of Monopolistic Competition*.
    It is also worth noting in this connection that all the customary measures of the degree of competition — demand elasticities, concentration ratios, Gini indexes, etc. — are not only independent, but designed to be independent, of scale.

true that man power is a singularly unfortunate index for measuring size when one is trying to explain differences in labor productivity,[4] since equality of size in terms of man power would imply, in terms of the more relevant index, output, a disparity in size between America and Europe of the order of two to one or even three to one. Nevertheless, I doubt if such disparities in size are the main explanation, especially when experience suggests that there are also other, equally valid and equally important explanations. For example, one would expect a plant *designed* to employ 500 workers to be more efficient than one that has *grown* to the same size (however measured) from small beginnings, by the addition here of a new workshop, there of a few machine tools — even if each and every addition was as efficient as could be. Indeed, what seems relevant from the point of view of technological efficiency is the scale, not of the plant, but of the individual investment or investments that created the plant. In other words, it is the firm's rate of growth, or rather the size of the discrete steps by which it grows that determines what method of production is chosen as being the most profitable one.

This suggests the growth rate of the economy or of the market in which the firm operates as an important factor — although this again is by no means the only or the main factor. For national differences in rates of growth would not be very important if each firm were restricted to its proportionate share of the market. In other words, whether the firm's market grew at 3 or at 10 per cent per annum would probably make little difference to its scope for modernization and the adoption of mass-production methods. More important is the firm's ability (and inducement) to expand at a faster rate than the market in which it operates, by arrogating to itself more than its proportionate share of the market's expansion and by encroaching upon its competitors' markets. This, however, depends on the nature and degree of competition, which determine the cost to the firm of expanding its market at a faster rate than the market as a whole expands; and the degree of competition and cost of expansion may well be significantly different on the two sides of the Atlantic.

In other words, when an entrepreneur invests in new productive capacity, the method of production that will appear the most profitable to him depends on the size of the market to which the new capacity will cater; and this in turn depends partly on the natural accretion to his market, partly on its further expansion through competition. It seemed right to define the natural accretion to the firm's

4. It is, of course, the only one available.

market as its proportionate share in the expansion of the market as a whole; and it must be noted that the absolute size of this proportionate share depends on the firm's size (in terms of output) as well as on the rate of growth of the whole market. Hence, with the size of the average firm larger in America than in Europe,[5] and with the rate of expansion of the American economy also greater, we already have two factors that would make for greater technological progress in America. In addition, I shall argue presently that competition is yet another factor biased in America's favor.

The meaning of competition in this context is slightly different from that customarily attributed to it. For one thing, competition in the above sense is very much a long-run concept, while the preoccupation with price competition in so much of the current literature naturally puts the emphasis on the short run. For another, the customary preoccupation with income distribution and static welfare concepts makes one think of competition as a restraining force, which limits the would-be monopolist's control over supply, influence on price, and ability to enjoy a quiet life; whereas here, competition means the entrepreneur's opportunity to encroach upon the potential[6] and actual markets of his competitors, and — if they have similar opportunities — his inducement to encroach upon their markets before they encroach upon his. Competition always has to do with one firm's encroachment upon another's market; but in much of the past literature attention is centered on the behavior of the firm whose market is being threatened or encroached upon; whereas here our interest is focused on the firm that does the encroaching and on the effects of its encroaching on its own investment opportunities and decisions.[7]

A third difference between competition in the two senses is the following. Most of the beneficial effects of competition on income distribution, resource allocation and managerial efficiency can be brought about by the mere threat (or rather fear) of market encroach-

---

5. The size of the firm is relevant also for another reason. Expansion usually goes hand-in-hand with the replacement of existing capacity; and part of the market for the new capacity is created by the scrapping of old capacity. Hence, the more old capacity the firm has for scrapping, the less expansion is needed for providing the new capacity with an adequate market.

6. By a firm's potential market we mean its proportionate share in the expansion of the whole market.

7. For a statement of this difference in a slightly different context, see M. Abramovitz, "Monopolistic Selling in a Changing Economy," this *Journal*, Feb. 1938.

ment; and this threat may forever remain a threat and need never be carried out. In the realm of current production, therefore, competition may enforce the firm's good behavior in a way not unlike that of the sword that hung above Damocles' head. By contrast, the technological progress of the firm, if conditional upon a substantial expansion of output, depends on the firm's really encroaching upon its competitors' potential or actual markets. These distinctions suggest that the competition relevant to the firm's investment decisions is essentially aggressive and may diminish both the number of competitors and the degree of competition. Indeed, it is closely related to the conception of competition as a struggle to establish a monopoly position.[8]

Returning now to the comparison between American and European economies, there is, unfortunately, no comparative study of degrees of competition and, to judge by the information contained or cited in the eight "country papers," none is possible at present. A comparison between the United States and the United Kingdom has been made and shows a higher concentration ratio in the latter than in the former;[9] but one cannot generalize from this to a higher degree of concentration in Europe as a whole.[1] In contrast, however, to the lack of statistical information, institutional information is available, and it suggests that there may be a significant difference between American and European industry, with the representative American businessman facing lower costs and lesser difficulties than his European counterpart when he tries to expand his market sufficiently to render modern mass-production methods profitable.

To begin with, the American system of distribution, whereby the individual manufacturer markets relatively few products through a large network of retailers, probably makes for lower costs of market expansion than the European system of marketing a large variety of products through few but exclusive retail outlets. (Indeed, the mere

8. Schumpeter assumed — unnecessarily, I think, and mainly for the sake of paradox — that the entrepreneur needs a monopoly position to render innovation worth his while; I am arguing that he needs a market that exceeds a certain minimum size, which is largely determined by the engineering characteristics of the new facility he contemplates installing, and which may or may not imply a monopoly position.

9. Cf. Gideon Rosenbluth: "Measures of Concentration," in *Business Concentration and Price Policy*, A Conference of the Universities-National Bureau Committee for Economic Research (Princeton: Princeton University Press, 1955).

1. Whether concentration ratios are a suitable index for measuring competition in the sense here used is another matter; but with this we need not be concerned here.

fact that marketing as an activity is so much more specialized and highly developed in America than in Europe suggests that market expansion may well be cheaper and more easily accomplished in America.) Secondly, the American consumer's preference and expenditure pattern is probably more flexible and more easily swayed than the European's, partly perhaps because he is less conservative in such matters, and partly because his higher standard of living gives him more scope for flexibility.[2] Thirdly, the scope for expansion by encroaching upon the markets of inefficient firms is probably greater in America than in Europe, where the inefficient firms are usually the small family businesses, whose staying power and ability to sustain losses is much enhanced by the fact that their value to their owners consists not only in the profits they yield but also in the independent position and secure employment opportunities they provide.[3] A fourth and related factor is that family business, still preponderant in Europe, probably has less easy access to the capital and entrepreneurial talent needed for expansion if the owners are determined, as they often are, to keep the business a strictly family affair.

Such and similar factors seem to be important determinants of the cost of expansion and hence of the firm's ability and inducement[4] to adopt new methods of production whose profitability hinges on a high rate of output. They do not seem to have been discussed at the conference; as indeed there seems to have been little explicit analysis of the influence of competition on the nature of the firm's investment. By implication, however, and in slightly different guises, the subject was very much in the foreground. To begin with, Professor Hennipman's discussion of innovations follows lines that tie in with those sketched above. Secondly, there is evident, in several of the papers, a general trend towards the use of long-term concepts in dealing with problems of competition. Professor Chamberlin warns against relying exclusively on the elasticity concept for "Measuring the Degree

2. The pronounced positive correlation between the variance and the mean of family expenditures does not prove, but suggests strongly, that there is such a difference. Cf. M. Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, Vol. 32, p. 675; D. S. Brady, "Variations in Family Living Expenditures," *op. cit.*, Vol. 33, p. 385; S. J. Prais & H. S. Houthakker, *The Analysis of Family Budgets* (Cambridge University Press, 1955), p. 56.

3. Cf. J. Steindl, *Small and Big Business* (Oxford: Blackwell, 1945), chap. 6, for an analysis of the small firm's staying power.

4. One entrepreneur's ability to adopt new methods of production is his competitors' inducement to do likewise.

of Monopoly and Competition"; and Professor Bain, in his very interesting paper on "Conditions of Entry and the Emergence of Monopoly," develops a complete long-run theory of price determination. He relates the firm's price not to its own costs but to the minimal average cost of the most-favored firm and explains the margin between long-run price and minimal average cost in terms of the cost of entry. These concepts, which resemble closely our cost of expansion and socially necessary minimum cost, are used by Bain solely to explain prices, profits, and profit margins; but their relevance for the firm's investment policy is obvious.

Thirdly, there was much talk at the conference, both in the papers presented and in the ensuing discussion, of the importance of distinguishing between different kinds of competition. Mr. Corwin Edwards, opening the conference with a statement of the "Issues in the Monopoly Problem," clearly defined a variety of meanings of the term competition. But the interest of the conference in these matters was aroused only much later, by which time Edwards' clear-cut definitions seem to have been forgotten and the discussion revolved around the more intriguing but less clearly defined concepts of active competition and workable competition. There seems to have been a consensus that these concepts are important; and that active or workable competition may be present or absent independently of whether competition in the traditional sense is present or absent. But this, unfortunately, is about as far as the subject was pursued at the conference. To judge by internal evidence, by the fine display of winemanship in the illustrative examples, and by my own independent knowledge of the mellow charm and three-star restaurant of Talloires, the conferees were in no mood to use their considerable analytic powers for a rigorous definition and analysis of these concepts. There are definitions, but none too clear. Sometimes active competition seems to mean the entrepreneur's ability to expand his market and thus render profitable the use of a better method of production — this meaning is implied, for example, in Professor Clark's interesting discussion of the time element in competition (pp. 326–28). At other times, however, active competition seems to mean, not the market conditions that render innovation profitable, but the entrepreneur's innovating activity itself. This latter, of course, is the *desideratum;* but the former concept is the tool needed for analyzing the conditions that hamper or promote the latter's advent.

It may seem unfair to the book under review to discuss at such

length a problem it fails to solve. There are, indeed, many other subjects fully and satisfactorily treated in the volume; but its main value, at least in this reviewer's opinion, is the background and stimulus it provides for the solution, by its readers, of some of the unsettled problems of monopoly and competition.

TIBOR SCITOVSKY

STANFORD UNIVERSITY

THE PROPORTIONALITY CONTROVERSY AND
THE THEORY OF PRODUCTION

By Harvey Leibenstein

In the theory of the firm much is made of the observation that certain factors are lumpy. In fact, many writers have gone so far as to attribute both increasing and decreasing returns solely to this phenomenon. Professor Chamberlin challenged this view in a well known article, but defenders of what may be called the "proportionalist position" were quick to reply.[1] (By the "proportionalist position" we refer to the notion that, given divisibility, the multiplication or division of all factors by a constant will yield proportional returns; in other words the idea that, given divisibility, production functions are necessarily homogeneous.)[2] A recent article by Messrs. Whitin and Peston reinforces the Chamberlinian view by describing "several situations where the rational entrepreneur will vary certain factors less than proportionately with output."[3] But in these discussions an important methodological point has been missed. In this note we shall explore this point and show how it can help to clarify some aspects of the problem.

One of the reasons why controversies on theoretical issues are often difficult to settle is that the theory, in its development towards greater and greater generality, becomes so terse that it does not permit certain significant distinctions to be made. As a result implicit assumptions go unrecognized and appear to lead inevitably to what are in fact wrong conclusions. We suggest that this is the case with the conventional theory of production of the firm. When the necessary distinctions are made, some of the central issues in the controversy are clarified, and we shall see that, once these distinctions are accepted, the "proportionalist position" becomes untenable.

1. E. H. Chamberlin, "Proportionality, Divisibility and Economies of Scale," this *Journal*, Feb. 1948, pp. 229–62; "Comments" by A. N. McLeod and F. H. Hahn, "Reply" by E. H. Chamberlin, *ibid.*, Feb. 1949, pp. 128–43. The first article is incorporated in *The Theory of Monopolistic Competition*, 6th ed., as Appendix B.
   2. What the outcome of the debate has been is not at all clear. If the writer can judge on the basis of conversations he has had with some economists on the problem, he must conclude that the "proportionalist position" is still held quite tenaciously in some quarters.
   3. T. M. Whitin and M. H. Peston, "Random Variations, Risk, and Returns to Scale," this *Journal*, Nov. 1954, pp. 603–12.

1. *The Issue.* According to the proportionalist argument if the factors of production are infinitely divisible then this inevitably leads to long-run constant costs. Since there is some optimum combination of the factors that will yield an output at lowest average unit cost, then, given infinite divisibility, it follows that we could obtain greater or lesser outputs at the same average cost simply by dividing or multiplying the optimum combination by an appropriate constant. At first blush this appears to be an eminently reasonable point of view. Yet it is quite incorrect.

2. *The Meaning of Factor Divisibility.* Before embarking on our analysis proper we have to indicate what is meant by a fractional unit of a factor. Assuming that all units are homogeneous, the question that arises is how to define half units where it is not meaningful to divide a unit physically. In this case we may, for the sake of argument, accept what appears to be the "proportionalist position" and define half units in terms of efficiency. Namely, one hundred half-men are equivalent to fifty whole men, and this equivalence is in terms of what they can do. Thus, the entrepreneur who hires one hundred half-men, is just as well off as he who hires fifty whole men or two hundred quarter-men, etc. In other words, our definition implies that *under similar circumstances* fractional units, no matter how finely divided, operate just as efficiently as whole units. Surely this meets the proportionalist ideal of perfect divisibility, and yet we shall see that proportional returns to scale do not follow.

Before presenting the elements of our model two aspects of the problem must be made explicit. First, it must be understood that the controversy is about the divisibility of *factors* and its consequences, and not about *other* aspects of the productive process. In other words, when writers on the subject speak of the divisibility of factors we assume that we are to take them literally and that they *are* speaking about factors of production. Second, we assume that the theory we are considering is not merely a formal matter but that it has to do with the productive process as it exists in the real world, although the theory may abstract from many aspects of that process.

3. *The Production Model.* The theory of production is in terms of two polar categories, inputs and outputs. While it is true that writers on production theory are aware that there is more to the process than merely combining inputs and obtaining outputs, and they sometimes allude to intervening elements, it nevertheless remains the case that the formal theory is usually presented in terms of inputs and outputs. For many problems the input-output approach is

sufficient. But with respect to the controversy under discussion a model in which the intervening categories are made explicit is the more adequate.

Recalling that the issue at stake is the contention that complete divisibility is a sufficient condition for constant long-run average costs, it follows that to disprove this proposition we need conceive of merely a single possible case in which this is not so. Thus, if we can visualize a model of the production process in which the proposition at issue need not hold, then we have proven our case.

And now to our model. The definitions of the basic categories of our model follow:

(1) A *commodity* is the entity that is the object of the productive process, and that has a specific set of attributes or specifications.

(2) An *activity* is our primitive concept. It refers to those necessary acts carried out by a factor, or functions of a factor, necessary in the productive process. We define a set of related activities as an *operation*.

(3) A *process* is any specific set of operations used to produce the commodity in question. There may be a number of possible alternative processes.

(4) A *factor* is an entity, units of which can be purchased on the market, that has the capacity to carry out one or more activities.

(5) By a *firm* we refer to the entity that purchases factors, creates commodities, and sells commodities.

Our vision of the productive process is this: The firm purchases all the necessary factors. The firm's purpose in doing so is to create some quantity of a commodity which can be defined in terms of a set of attributes. The attributes are created by a set of operations. We refer to any specific set of operations as a process. The operations are performed by or made up of activities. Our notion of an activity is a rather broad one so that it contains any contribution to the creation of the commodity. For example, we conceive of some specific raw material, say a piece of wood, as capable of performing the activity of becoming the necessary physical matter out of which the commodity, say a bench, is made.

Since we assume perfect divisibility, the firm can purchase its factors in any conceivable combination. For some given output the commodity is produced at minimum costs for that output and with a specific ratio of the factors. The question is whether the firm can also produce any other output at the same cost per unit if the factors are purchased in exactly the same proportion as the alternative out-

puts. The answer depends on whether the quantity of any activity that has to be carried on is a function of the output of the commodity. Now clearly in any conceivable productive process *not all activities or operations need be functions of output.* Indeed, it is unlikely that there is any modern productive process in which *all* the activities or operations depend entirely on the quantity produced. Once we accept this possibility then we can quite readily see that perfect divisibility of *factors* (provided we accept the distinctions between factors, activities, and operations) is not a sufficient condition for long-run constant costs.

The number of times any activity has to be performed may vary directly with the amount of any given factor that is hired, the variety of the factors, the number of operations, or the number of other activities. The relationship need not be a proportional one. For example, the activities of starting and stopping may be functions of the number and variety of operations that are performed. Nor need we assume that any particular activity need always be carried on for any particular output. Some activities may have to be carried on regardless of output while other activities and operations may be required when some other variable in the production process reaches a certain minimum size, while still others may grow or decline in a nonproportional fashion to other variables. The extent to which the condition just alluded to holds depends on the nature of the productive process.

The critical reader may assert that all we have done is simply to shift our grounds from lumpy factors to "lumpy activities" and operations. But this is not the point. For what we come out with is that there is something inherent in the productive process and not in the divisibility of factors that is the cause of nonhomogeneous production functions.

Furthermore, we have not really shifted ground by changing the definition of a factor, for the issue is one of *meaningfulness* and not of definition. There is meaning to a "lumpy factor" but there is *no* meaning to a "lumpy activity" or to a "divisible activity." Consider, for example, the production process of publishing a book. One of the activities is proofreading. Once it is done it need be done no more, regardless of the number printed, the thickness of the pages, or any other quality of the book. It is meaningless to say that proofreading is a lumpy or indivisible activity. It is simply an activity determined by the technical requirements of producing a book. The fact that it has to be done only once irrespective of the number of

copies produced is entirely unrelated to the bulkiness or divisibility of the factors of production employed. It is just as meaningless to half proofread something as it is to half kick a ball. You either kick it or you do not. An activity is either performed or it is not. The fact of its performance has nothing to do with the size of the ball, the size of the kicker, or their divisibility or lack of it. Something is not proofread if only half the pages are read. In a similar vein to "half-switch" on a light is a meaningless idea. Without belaboring this notion further it should be clear that some combinations of words simply have no meaning and it is a methodological error to act as though they had. By making inadequate distinctions, or by leaving certain elements implicit, some writers seem not to have realized that they were indirectly advocating combinations of ideas that really make no sense.

An objection that may be raised is that our definition of a factor involves an unnecessary distinction; that the firm really purchases the activities performed, and therefore, the activities and the factor are one and the same thing.

In reply we argue that it is surely legitimate to distinguish what is in fact distinguishable. The factors of production that are hired in the market are in fact not the same thing as the uses to which they are put. The factors may be divisible and yet the activities that they perform are discreet. It may take multiple (or fractional) units of a factor to carry out an activity. Under such circumstances we do not violate the assumption of the divisibility of a factor.

But even if we think of factors solely in terms of the bundle of activities performed it does not change the basic argument. The crucial distinction to be recognized is that the nature of the factors supplied must be separated conceptually from the nature of the activities *needed* in the production process. The nature of the factors supplied is determined by conditions outside the production method under consideration. For example, the attributes and specifications of the commodity need not, and do not, determine the characteristics of the factors. But, the attributes and specifications of the commodity do determine, in part at least, the activities and operations necessary for the production of that commodity. That is, the activities and operations required in the production process are determined, to some extent, by technical considerations. Furthermore, the technical considerations may be such that they require operations and activities to be performed that are nonproportional to output.

In order to sharpen the argument it may be useful to classify

activities into three categories. (1) Direct proportional activities: those activities which are involved in operations on the commodity in question, and where the quantity of the particular activity is proportional to output. (2) Direct nonproportional activities: these are activities that are directly connected with producing the commodity but where the amount of the activity is not proportional to output. (3) Indirect activities: here we have in mind activities such as personnel administration, record keeping, and others that do not involve contact with the commodity but are necessary to the operation of the firm.

Now the implicit assumption of the "proportionalist position" must be that if factors were completely divisible then *all* activities would be direct proportional activities. This implies that all direct nonproportional activities and all indirect activities carried on by firms at the present time are entirely due to the indivisibilities of factors. But this last implication of the "proportionalist position" is most difficult to believe if we consider some of the specific nonproportional activities that have to be carried on by firms. For example, such activities as production planning, the issuance of orders to personnel, keeping records, the channelling and collecting of information internal to the firm, the channelling and collecting of information external to the firm, the administration of personnel, the administration of credit and finance, as well as a host of similar control activities need not be, and probably are not, directly proportional to output. Many of these activities have to be carried on regardless of output and one cannot see how the fact that factors were divisible would either eliminate their need or transform them to proportional activities. Consider factors that have some degree of durability. Surely the appropriate depreciation formula would not always be of the form $D = NQ$, where D is the amount of depreciation and Q the quantity produced. The "proportionalist" view would imply not only that all depreciation costs are user costs, but also that user costs are always proportional to output if factors are divisible. But the fact of the matter is that the physical nature of durable goods is not of this kind. Some depreciation takes place for reasons other than use. The case is especially clear with respect to obsolescence. The risks of obsolescence, and the fact of obsolescence, surely do not depend on the divisibility of factors. Another interesting possibility is a chemical process where a catalytic agent is involved. Again, the amount of the catalytic agent need not be proportional to output. This is a matter of the laws of chemistry and not of economics or of

the divisibility of factors. Other interesting examples in this vein, especially for cases where risk and random variations are involved, are mentioned in the article by Whitin and Peston.[4]

In sum our argument boils down to the fact that many production processes necessitate indirect activities and operations, as well as direct nonproportional activities. The issue at stake is not a matter of definition. Essentially, it is both an empirical matter and a question about the meaningfulness of ideas. The empirical aspect depends on what the firm has to do in order to produce a commodity and maintain the firm as a continuing entity, and this does not depend entirely on the divisibility of factors. It does depend on the objective nature of commodities and production activities, and on the technical requirements of production processes, which in turn depend in part on the laws of physics and chemistry and on the arts of engineering and administration.

The lack of meaningfulness of some of the assertions about the consequences of factor divisibility have not been recognized heretofore because they were based on a theory in which there are inadequate distinctions. As a result it was not recognized that certain implicit elements in the argument implied combinations of ideas that were in essence nonsensical. By breaking down the theory into more elements we are enabled to make these necessary distinctions and thus bring to the fore the meaningless nature of some of the arguments. Specifically, we have seen that while it may often be sensible to speak of the divisibility of the factors it is quite meaningless to think even implicitly in terms of divisible activities. And since activities need not be proportional to output if factors are divisible, then factor divisibility need not lead to constant returns to scale.

HARVEY LEIBENSTEIN.

UNIVERSITY OF CALIFORNIA,
BERKELEY

4. Whitin and Peston, *op. cit.*

WALRAS, LEONTIEF, AND THE INTERDEPENDENCE OF ECONOMIC ACTIVITIES: COMMENT

By James M. Henderson and Richard E. Quandt

In a recent article,[1] Mr. Robert Kuenne presented a revised version of Walras' general equilibrium system.[2] He devotes most of his attention to the introduction of money into the Walrasian system, and we shall confine our comments to this problem. We shall distinguish three types of general equilibrium systems: (1) a *static* system in which positive stocks of paper money are inconsistent with equilibrium, (2) an *imperfectly dynamic* system in which money is implicitly assigned the credit function, but in which positive money stocks are still inconsistent with equilibrium, and (3) a *dynamic* system in which positive money stocks are consistent with equilibrium. We intend to demonstrate that both the original Walrasian system after his introduction of money and the Kuenne revision are *imperfectly dynamic* systems, and that the Kuenne revision is subject to the basic criticism that can be leveled against the Walrasian system and to additional criticism as well.

Before he attempted to introduce money Walras had constructed a static general equilibrium system in which consumers maximize timeless utility functions subject to timeless budget restraints. Consumers sell factor services to entrepreneurs who in turn convert these into commodities which they sell to consumers. Consumers' expenditures must equal their receipts from the sale of factor services. From the maximizing procedure Walras derives the well-known equilibrium conditions that the ratio of the marginal utilities of any two commodities is equal to their price ratio. He requires that in equilibrium the supply of every commodity be equal to the demand for it. The quantities of the commodities produced, together with their fixed production coefficients, determine the demands for factor services by

1. "Walras, Leontief, and the Interdependence of Economic Activities," this *Journal*, LXVIII, (1954), 323–54.

2. Walras' system is given in his *Éléments d'Économie Politique Pure*. We shall refer both to the French edition (ed. déf., Paris, 1926), and the recent English translation by William Jaffé, *Elements of Pure Economics* (London, 1954). The first reference is to the French edition, and the second in parenthesis to the English translation.

entrepreneurs which are required to equal the supplies offered by consumers. After imposing the equilibrium condition that the price of each commodity be equal to its cost, he is able to determine the outputs and relative prices of the system.

Finite money prices are inconsistent with positive money stocks in this system.[3] Money does not enter the utility functions of the consumers. If a consumer has an initial money stock from which he derives no utility, he will attempt to exchange it for commodities from which he derives utility. Similarly, the entrepreneur will not desire to hold money in this timeless system. Since all consumers and entrepreneurs will attempt to exchange their money stocks for commodities and factor services, the price of money in terms of commodities will tend towards zero. Positive money stocks are clearly inconsistent with the maximizing behavior of consumers and entrepreneurs. There can only be a money of account, and absolute prices can be determined by setting the price of some arbitrary commodity (the *numéraire*) equal to unity.

Walras introduced money as a form of circulating capital:

"Production . . . requires a certain lapse of time. We shall resolve the . . . difficulty purely and simply by ignoring the time element at this point. And later on . . . we shall bring in *circulating capital* and *money* and thereby make it possible for production services to be transformed into products instantaneously, provided that the consumers pay the interest charges on the capital required for this sort of transformation.[4]

He assumed that consumers desire a service of availability, or an inventory service, of the goods they consume, and that they desire this service both *in kind* and *in money*. Thus, consumers derive a certain satisfaction from the possession of money and their stock of money held for inventory purposes enters their utility functions in an indirect way. The amount of money that consumers wish to hold in the form of "postponable inventory services" is determined by the condition of the proportionality of marginal utilities and prices. The price of these money services is equal to the price of money in terms

3. This proposition was stated by Don Patinkin, "Relative Prices, Say's Law and the Demand for Money," *Econometrica*, XVI (1948), and has been the subject of extensive discussion. The definitive treatment was given by Jacob Marschak, "The Rationale of the Demand for Money and of 'Money Illusion,'" Cowles Commission Paper, New Series No. 46, (1951). Also see G. S. Becker and W. J. Baumol, "The Classical Monetary Theory: The Outcome of the Discussion," *Economica*, N.S. XIX (1952). Later we shall indicate how this static system must be modified to allow the introduction of money. Here we only state that money cannot be introduced unless the system is modified.

4. Walras, p. 215 (242). Jaffé's translation and Walras' italics. Walras' monetary analysis is contained in Lesson 29, pp. 297–308 (315–24).

of the *numéraire* multiplied by the interest rate. This modification
of the static system is accomplished by the introduction of time since
the desire for inventories presupposes some process *in time*, but his
analysis must be classified as *imperfectly dynamic* because time does
not enter into the explicit formulation of the model; the economic
units' horizon is still limited to only one instant, although implicitly
the system extends over a time-period.

Kuenne rejects the postponable inventory service approach and
looks upon money as " . . . a flywheel converting discontinuous
income receipts into a smooth path of consumption through time. . . ."[5]
He introduces money on the consumers' side by means of a disaggre-
gated Marshallian-*k* analysis, and states the equation for the aggre-
gate offer of money by consumers:[6]

$$\overline{X}_u = Q_u - \frac{\Sigma K_{y_k} X_{y_k} p_{y_k} + K_e X_e p_e}{p_u}$$

where $\overline{X}_u$ is the aggregate offer of money, Q_u is the original stock of
money held entirely by consumers, K_{yk} is the proportion of their
expenditure on the *k*th commodity which consumers desire to hold
in the form of money, X_{y_k} is the total demand of consumers for the
*k*th commodity, and p_{y_k} is the price of the *k*th commodity in terms
of the *numéraire*. K_e, X_e and p_e have similar interpretations with
regard to the consumers' purchases of securities. Kuenne's innova-
tion consists of substituting institutionally determined *k*'s for Walras'
marginally determined holdings of money and thereby completely
divorcing money from the utility functions.

Kuenne's modifications of the Walrasian analysis are more
modest on the demand side of money. He formulates an equation
expressing the demand for money by entrepreneurs, and requires
this demand to equal the supply offered by consumers:[7]

$$\overline{X}_u = \Sigma a_{uy_k} \overline{X}_{y_k} + \Sigma a_{uz_j} \overline{X}_{z_j}$$

where a_{uy_k} is a coefficient expressing the average amount of money

5. Kuenne, *op. cit.*, p. 332.
6. *Ibid.*, p. 333. Kuenne's notation is retained. Kuenne derives this equa-
tion by aggregating the offers of money by the individual consumers:

$$\overline{X}_{iu} = q_{iu} - \frac{\Sigma k_{iy_k} X_{iy_k} p_{y_k} + k_{ie} X_{ie} p_e}{p_u}$$

where *i* refers to the *i*th individual. The relationship between the individual
k_{iy_k} and the aggregate K_{y_k} is not apparent. Kuenne must have assumed that the
k's are the same for every individual.
7. *Ibid.*, p. 337.

held by entrepreneurs per unit of output of the kth commodity, a_{uz_j} per unit of output of the jth capital good, $\overline{X}_{y_k}$ the output of the kth commodity, $\overline{X}_{z_j}$ the output of the jth capital good. Walras treated money coefficients as fixed but Kuenne allows them to vary.[8]

Both Walras and Kuenne have taken the static system, placed it within the framework of a single time-period and introduced money to synchronize receipts and expenditures during this period. But the function of money is not identical in the Walrasian analysis and in the Kuenne version. Walras treats money as providing postponable inventory services and therefore as a form of circulating capital. Kuenne uses money for the purpose of allowing the consumer to bridge the gap between income receipts and expenditures. Kuenne, like Walras, introduces a dynamic element into the system, but retains the timeless equations of Walras' static system. The consumers' or entrepreneurs' planning is not allowed to extend over more than a single instant. The monetary analysis grafted on to the static system is clearly inconsistent with the static equations.

Let us assume that time enters the utility functions of the consumers so that the utility of a commodity is dependent upon the moment at which it is purchased during the period, and that time is introduced into the production functions.[9] We can now postulate that consumers desire to maximize their utility functions for a given time period. Following Kuenne's assumption, consumers receive their income discontinuously, and production and consumption take place continuously during the period. If consumers receive their incomes before they consume, they are financing the entrepreneurs, or if they receive their incomes after they consume, the entrepreneurs are financing them. In either case, the net deficit of one group equals the net surplus of the other. At the end of the period consumers' income equals consumers' expenditures, and entrepreneurs' expenditures equal their receipts, and although one group

8. Kuenne assumes the money coefficients to be jointly determined by institutional factors and prices (p. 338). This is a strange condition since one would expect something to be either institutionally determined, or price-determined, but not both. The only reasonable way of interpreting this condition seems to be to assume that entrepreneurs desire to hold a cash balance which is a certain proportion (institutionally determined) of their expenditures on factor services (prices multiplied by quantities). If this is the case, the entrepreneurs' cash balances are determined in the same way as the consumers' cash balances.

9. Both Walras and Kuenne implicitly assume that time enters into the utility and production functions, but we cannot interpret their implicit analysis to allow for more than a single period.

finances the other during the period, neither group is a creditor or debtor at the end of the period.

The function of financing in this case is a function of credit, not a function of money. This can be illustrated by first showing how credit can perform this function, and then showing why money cannot. Debts are in existence during the period, but are zero at the end. Since credit performs the function of financing, it must enter the utility functions of the consumers. If they are debtors, credit allows them to consume when they derive the greatest amount of utility subject to the restrictions of the model, and on the margin they balance the interest charge for credit against the additional utility gained through its use. If they are creditors, they must balance the delay of consumption against the interest earned. Synchronization can take place through the use of credit, and money is not necessary to perform this function. We need only a unit of account in which to express prices and debts.

We have assumed no impediments to borrowing and lending. Hicks has suggested that money can only be introduced if there is some impediment to perfect borrowing and lending.[1] Walras did not specify such impediments. Kuenne discusses Hicks's suggestion, but fails to make such impediments an essential part of his analysis.[2] Under these conditions the "postponable inventory services" of Walras cause credit and not money to enter into the utility function. Therefore, if borrowing and lending are possible without impediments, no money stocks will be held at any time.

Kuenne's analysis contains an additional inconsistency. First, his assumption that money enables the consumer to spread his purchases over time suggests that money enters the utility function. However, this contradicts the institutional determinism implied by his version of the Cambridge equation. If he allows money to enter the utility functions, he can derive demand and supply equations for money from the utility functions and the budget restraints. Then the Cambridge equation, showing the consumers' supply of money, is either redundant or inconsistent with the demand and supply functions derived from the utility functions. In addition, it has been shown that the assumption that money enters into the utility function is neither sufficient to explain positive money stocks if unlimited borrowing is permitted, nor necessary if credit enters the utility function.

1. J. R. Hicks, *Value and Capital*, chap. XIX.
2. Kuenne, pp. 328–29.

These considerations indicate the manner in which the Walrasian system and the Kuenne version would have to be extended and transformed in order to provide a consistent explanation of the function of money in general equilibrium. It is necessary to dynamize the equations of the model in order to provide such an explanation.[3] Time and expectations can be introduced into the utility functions as soon as we allow the model to extend over the individuals' horizon. Among the sufficient conditions for the effective introduction of money in a dynamic system are uncertainty, impediments to borrowing and lending, and the expectation by at least one individual of falling interest rates.

James M. Henderson.
Richard E. Quandt.

Harvard University

REPLY

By Robert E. Kuenne

I shall begin my response to the criticisms of Messrs. Henderson and Quandt by defining those areas in which I feel no disagreement exists. The page numbers in parentheses refer to the location of relevant points in my original article.

First, we seem to agree that money stocks are inconsistent with a static system which has no implicit time dimension. As I shall indicate below, I feel that this proposition does not say much, but it can be looked upon as the starting point for both of our analyses (334–35).

Second, as I think my original article made clear, I should certainly agree that if aspects which Henderson and Quandt denote as "dynamic" are introduced into Walras' system, the difficulties of integrating the monetary equation into the "real" system in a realistic way — they speak of making the two "consistent" — disappear (328, n. 4). Moreover, it is agreed that *only* if this is done can a

---

3. By "dynamic" we mean "multiperiod" in the Hicksian sense, but the present definition of "dynamic" is not essential for the argument. However, the models we have classified as imperfectly dynamic must be modified to allow the introduction of money.

realistic (as distinguished from a determinate) theory of money be obtained (334, 336, 347).[1]

If this is a fair statement it is to the role of money in the system which they have called "imperfectly dynamic" that we must look for the substantive issues which divide us. To see these more clearly, let us deal summarily with two ambiguities in my presentation which they criticize justifiably, but which I feel they would agree are of no real importance to the argument.

First, in my treatment of the supply of money, the individuals' equations containing institutional constants were combined into an aggregate function with aggregative constants whose relationship to the individuals' constants was not made clear. This conversion was a last minute change resulting from my dissatisfaction with the retention of the individual demand and supply functions — the only satisfactory way of coping with individual differences. They indicate the correct interpretation these aggregative constants must receive (333).

Second, in dealing with the entrepreneurs' demand for money I referred to these needs as being determined essentially by institutions "modified by the indirect influence of prices" (338). The phrase was an unhappy one, as they point out. At this stage of the analysis I was interested in pointing up the peculiar relations between money inputs and "substitutional" inputs which were to be reflected in the equilibrium conditions for this sector; the indirect influence of prices referred to was meant to reflect the result of a change in the price of (say) labor upon the quantity of all factors used and the effect of these quantitative changes upon the demand for money. It was explained immediately below this statement that such needs varied with average expenditures on inputs per unit of product and that the treatment was analogous to that for households (339). However, unless I fail to see correctly, this is only a quarrel over words and I plead guilty of clumsy phraseology.

Now, let us deal with the authors' criticisms of Walras' monetary theory. Their basic argument is this: suppose time lags between production and consumption are introduced implicitly into the static model. Suppose also that riskless borrowing and lending may be carried on without impediment, let us say by allowing all economic units to issue and accept promissory notes or open-book credit for all purchases and sales. It is such credit, not money, which enters the

---

1. By a determinate theory of money I mean one in which the price of money is positive and therefore the absolute price level noninfinite.

utility function, and it obviates the need for stocks of money. Only if impediments to perfect credit facilities are admitted to the model does the *possibility* of money stocks arise.

I thought I had dealt adequately with this problem in so far as I (1) discussed Hicks' frictions as my critics note, (2) quoted from Marget exactly the same point that they make, and (3) ended with the conclusion that some such impediment had to be introduced to provide a *raison d'être* for cash balances (329). For example, the cost and trouble of bookkeeping or of settlement at the end of the period for large numbers of transactions is sufficient reason to expect the emergence of an impersonal synchronizing instrument called money.

Henderson and Quandt acknowledge that my article discussed the point although they do not mention that I specifically assumed such impediment. They say that I fail to make such impediments an "essential" part of my analysis. By assuming the existence of money in the Walrasian framework, I submit, such impediments become an inseparable part of the analysis through providing the *raison d'être* for balances referred to in the article (329). Would the authors require that the costs to individuals of transacting by credit be built into the system before such assumptions become integral? How, may I ask, did Hicks make such impediments an essential part of his stationary state analysis other than by *assuming* that money had a utility by virtue of impediments in the lending process, and that bills therefore stood at a determinate discount to it?

In any event, the authors accept the possibility of cash balances in an "imperfectly dynamic" system. But an observation which they make in note 9 forms a second objection to the existence of money stocks, in so far as they cannot interpret this "implicit analysis to allow for more than a single period." Evidently generosity has its limits. As I have argued in the article, money in the Walrasian schema is merely another type of asset yielding saleable services. In the absence of tomorrow consumers would refuse to accumulate stocks of all kinds: in other words, this interpretation would require that society be liquidated at the end of the current period, on the basis that Walras' analysis did not include future periods explicitly. Having admitted Walras' implicit inclusion of time *re* the current period, they "cannot" allow tomorrow to exist implicitly.

In my own criticism of Walras' failure to develop an explicit multi-period analysis, I said:

"Another failure of Walras' is more important for our purposes and cannot be corrected by a mere addendum. This flaw is his failure to allow the plans of

individuals to extend over more than one period (*except under the implicit stationary assumptions*). (345, italics supplied.)

My critics' position on this point seems extremely narrow to me. The explicit inclusion of the future to obtain the stationary state — a special case of Hicks's intertemporal analysis — changes nothing and avoids the need to liquidate society. Having swallowed the camel — an implicit present period — why do they strain at the gnat?

In summary, then, I feel that their basic case against Walras is weak. They argue (1) that if no impediments to perfect credit facilities exist the price of money will be zero, but that if such impediments do exist money can be held in Walras' imperfectly dynamic analysis; and (2) by implication, since an imperfectly dynamic system can admit today but not tomorrow (even if it be a certain, changeless tomorrow), no stocks of any kind will be held at the end of the period. I feel both arguments have been answered above.

My own critique of Walras' monetary theory hinged on two rather technical points: (1) that Patinkin demonstrated that individual excess supplies and demands for real goods could arise from proportional changes in all prices but that of money, given the income effect introduced by borrowing and lending, and therefore that the price of money was determinate at a nonzero level, and (2) that a Marshallian-k approach did not affect this determinacy and was preferable to the Walrasian method. My analysis under (1) was not attacked directly by the authors, and such indirect attacks as have been made have been dealt with above. Their criticism of my analysis centers about the second aspect.

Most of their attack upon me depends explicitly or implicitly upon the proposition that, given a utility for money and given a reason for not destroying stocks at the end of the period, the services of money *must* enter the utility function directly or indirectly. For example, they write "Kuenne's analysis contains an additional inconsistency. First, his assumption that money enables the consumer to spread his purchases over time suggests that money enters the utility function. . . ." Also, "Money does not enter the utility functions of consumers (in a static system). If a consumer has an initial money stock from which he derives no utility, he will attempt to exchange it for commodities from which he derives utility. . . ."

This narrow use of mathematics has led to much misunderstanding. For example, it is too easy to argue that in the classical analysis money did not enter the utility function, therefore money had no utility, therefore consumers had no reason to hold it, therefore

its price would be zero in a static economy. Let us admit more charitably that the classical economists were assuming that some reason for holding money existed, and therefore that it had utility, and that since it could have no utility in a static system they were not dealing with a static system. And this without the aid of mathematizers.

The essential point is that my removal of money from the utility function is not a denial that money has a utility for consumers. Nor, conversely, does the existence of such utility of money *necessarily* imply that it must enter the utility function used in the analysis. The utility function Walras used is merely a step to the derivation of marginal functions whose purpose is to allocate in an analytically meaningful way to a bundle of goods a total of satisfaction which in its essence is unassignable to specific goods in the bundle. It is a *methodological* decision as to whether the consumer's demand may best be treated on a "marginal" basis or some other basis — say, institutional." In some transcendental sense a unified function exists, but this is not the analytical utility function: this latter is defined for methodological convenience. As Patinkin has shown, given the desire to hold money, the determinacy of an absolute price level does not require money to enter the analytical utility function.[2]

After all, whole systems have been built which have not found the utility function useful. For example, Keynes did not wholly escape the allocation problem by aggregating consumption goods: savings exist as an active alternative. Yet I venture that in view of the host of difficulties incumbent upon the concept of a community utility surface, few would challenge his methodological right to determine consumption "institutionally" from the viewpoint of the present disagreement, even though such goods possess "utility."

The strange aspect of the arguments of those who hold these views is that they do not require that money enter the production function, implicitly accepting a nonmarginal approach to the determination of entrepreneurial cash balances. If, in the quotation from my critics above I substituted the words "entrepreneur" and "production function" for "consumer" and "utility function," would they still object to an "institutional" treatment?

Both Walras' and my own treatment assume that at a given *point* in time ("market day") a "desired" level of cash balances exists

2. My critics cite Patinkin's early argument which implies the necessity of money entering the utility function, a position he abandoned in a later article, "A Reconsideration of the General Equilibrium Theory of Money," *Review of Economic Studies*, XVIII (1950–1951), quoted on my page 336.

with which economic units begin and end the "week," but which may vary at any point in the week from the actual balances held. Also, this desired level is determined by the expected transactions to be made during the week. I think my own variation makes the concept of a "desired average cash balance" acceptable, avoids an excessive marginalism, and fits with no difficulty into an "imperfectly dynamic" system in a circular flow (which is all that my original article tried to show). Although the introduction of time, uncertainty, and expectations is necessary for a realistic theory, under these limiting assumptions a determinate theory of money is possible.

Lastly, I do not see how it is that my alteration means that the "function of money is not identical in the Walrasian analysis and in the Kuenne version." It seems to me that the manner in which cash balances are determined in my analysis does not alter the function of money, — it remains a form of circulating capital. This is doubly cloudy because in the authors' discussion of my theory they pointedly omit reference to the entrepreneurial demand for money. I can only say that I do not understand them on this matter.

ROBERT E. KUENNE.

UNIVERSITY OF VIRGINIA

PROFESSOR HANSEN
AND KEYNESIAN INTEREST THEORY

By Edward Nevin

According to Professor Hansen, Keynes's theory of the rate of interest is indeterminate. His argument is that Keynes alleged that the classical interest theory was indeterminate because the schedule relating savings to the rate of interest would shift as income changed, and incomes changed as the rate of interest changed. Exactly the same criticism, says Professor Hansen, is applicable to Keynesian interest theory, since the transactions demand for cash varies with the level of income:

According to the Keynesian theory the rate of interest is determined by the intersection of the supply schedule of money (perhaps interest-inelastic, if rigorously fixed by the monetary authority) and the demand schedule for money (the liquidity preference schedule). This analysis also is indeterminate because the liquidity preference schedule will shift up or down with changes in the income level. . . . In the Keynesian case the money supply and the demand schedules cannot give the rate of interest unless we already know the income level; in the classical case the demand and supply schedules for saving offer no solution until the income is known. Keynes' criticism of the classical theory applies equally to his own theory.[1]

This contention seems to rest on a rather rigorous interpretation of the concept of indeterminacy. Keynes did not convict the classical scheme of indeterminacy simply on the ground that a change in the rate of interest implied a change in the rate of investment, and therefore in incomes and savings at any given rate of interest. If economic theories became indeterminate every time two of their variables reacted upon each other precious few theories would escape the judgment. Keynes's point was not that the savings schedule shifted when the investment schedule shifted; the point was that they shifted *to exactly the same degree:*

The traditional analysis has been aware that saving depends on income but it has overlooked the fact that income depends on investment, in such fashion that,

1. Hansen, *A Guide to Keynes* (McGraw-Hill, 1953), chap. 7, pp. 140–41. The argument is substantially the same as that contained in Hansen's article, "Classical, Loanable-Fund and Keynesian Interest Theories," this *Journal*, LXV (Aug. 1951), p. 429.

when investment changes, income must necessarily change in just that degree which is necessary to make the change in saving equal to the change in investment.[2]
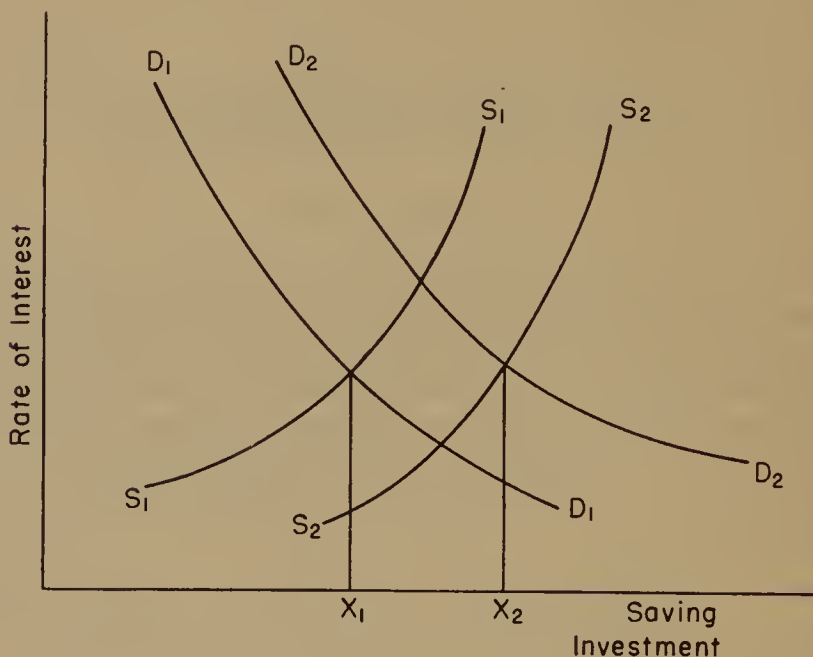

FIGURE I

This, of course, makes a great deal of difference.[3] The fact that when the DD' curve (Fig. 1) moved by, say, X_1X_2 the SS' curve *also* shifted X_1X_2 meant that the rate of interest was (directly) completely unaffected by the movement; to discover what *determined* the rate of interest one had to look elsewhere.

Now clearly the Keynesian theory is not indeterminate in this fundamental sense. The transactions demand may well vary with changes in incomes, but there is no reason whatever to suppose that a fall in the rate of interest, for example, would so raise the transactions demand through its effects on incomes that the rate of interest would be driven back to its original position. Such a sequence is, of course, impossible. If the rate of interest were unchanged, incomes

2. Keynes, *The General Theory of Employment, Interest and Money*, chap. 14, p. 184.

3. To use an analogy from algebra, the fact that certain unknowns vary together does not generally speaking make a system indeterminate if we have as many equations as unknowns. But if the unknowns maintain some constant relationship towards one another the equations cease to be independent (i.e., none indicates a new relationship between the unknowns which is not already implicit in the others) and the system is insoluble no matter how many equations there are. It is in this sense that Keynes argued that the classical theory of interest was indeterminate.

would be back at their original position and the transactions demand would not have increased. Similarly, a shift in the marginal efficiency schedule could not change the transactions demand through its effect on incomes in such a way that the rate of interest would be driven up and the rate of investment unaltered; such a sequence is internally inconsistent. In the Keynesian case, in other words, there is no question of a change in incomes causing two determinants to shift *equally*, and it is on this ground that the classical system was indeterminate.[4]

Hansen's point, in other words, is that there is not a unique liquidity-preference curve but a family of such curves, one for each income level. If this is sufficient to make the Keynesian theory indeterminate, it is clear that Hansen's procedure for solving the impasse leads us even more deeply into the morass than Keynes does. Hansen's solution is to resort to Hicksian *IS* and *LM* curves.[5] The *IS* curves determine the level of income associated with each rate of interest; the *LM* curves — given the state of liquidity preference
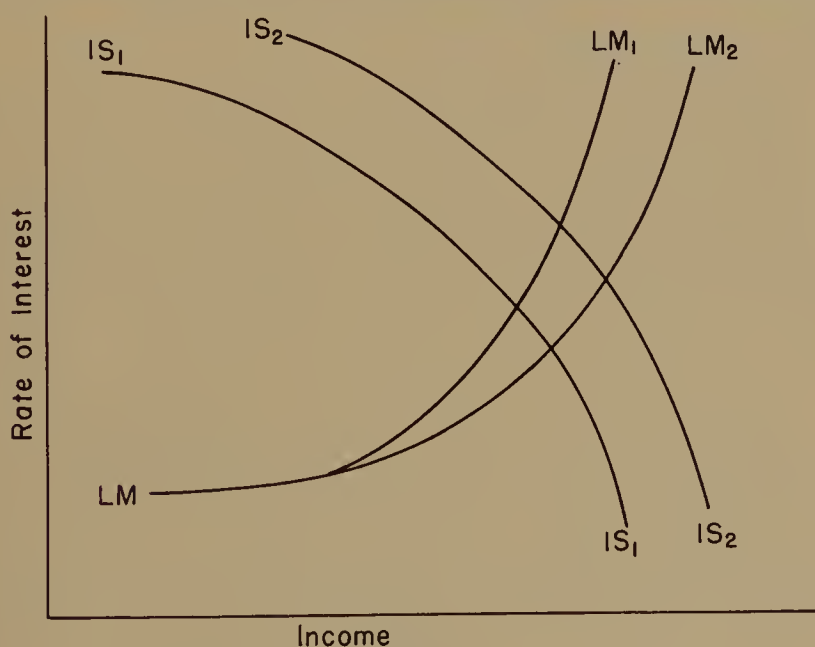
FIGURE II

and the volume of money — determine the rate of interest associated with each level of income:

4. The substance of the point made in the preceding paragraphs is made by R. F. Kahn, "Some Notes on Liquidity Preference," *The Manchester School*, Sept. 1954, p. 257.
5. Hansen, *op. cit.*, chap. 7, pp. 143–47.

Thus a determinate theory of interest is based on (1) the investment-demand function, (2) the saving function (or conversely the consumption function), (3) the liquidity preference function, and (4) the quantity of money.[6]

But Hansen's more complex formulation is equally open to the objection he raises against the Keynesian theory. As he himself suggests,[7] it is at least arguable that the investment-demand schedule is partly dependent on the level of incomes, so that a family of *IS* curves, one for each income level, may be more appropriate than a single schedule. Does this not land us back where we started? The *IS* curve is supposed to relate the rate of interest to the level of income, but until we know what the level of income is how do we know which *IS* curve is relevant? Nor is this all. The slope of the *IS* curve depends upon the saving function. But the saving function depends on the level of income; until we know the level of income, therefore, we do not know the slope of the *IS* curve. Neither the slope nor the position of the *IS* curve is known, therefore, until the level of income is known; we cannot tell where the *IS* and *LM* curves intersect — i.e., we cannot know the rate of interest — until we first know the level of income.

The difference between Keynes's formulation and that suggested by Hansen, then, is not that one is determinate — in Hansen's sense — and the other is not. The only difference is that the Keynesian *LM* curve relates the rate of interest to the quantity of money, incomes assumed constant, while the Hicksian *LM* curve relates the rate of interest to the level of income, the quantity of money assumed constant. The one formulation requires a separate curve for each level of income, the other a separate curve for each quantity of money.

One's preference as between these two types of formulation will be conditioned largely by the kind of problem which is being handled. It is perhaps relevant to recall that the Keynesian structure was one of equilibrium analysis — "the lagging adjustment that the economic system makes in response to the introduction of a disturbance is indeed often skipped over, and attention is directed to the equilibrium (or normal) magnitudes and relationships of the relevant variables."[8] In such an analysis, income is constant at the moment of examination, and the object will be primarily to consider the impact on the system of those variables which are most likely to change. It would be more helpful, perhaps, to assume income constant and consider the immedi-

6. Hansen, *op. cit.*, p. 146.
7. *Ibid.*, p. 143, n. 2.
8. Hansen, *op. cit.*, chap. 2, p. 47.

ate effect of variations in the money supply, rather than to treat the money supply as fixed and consider the repercussions of income changes on the rate of interest *after* the system has moved away from its original equilibrium. But that is a matter on which views will vary; perhaps it is just as well for the development of our theoretical equipment. It may be, in Keynes's own words, "important that opinions should differ."[9]

EDWARD NEVIN.

UNIVERSITY COLLEGE OF WALES
ABERYSTWYTH

COMMENT

By ALVIN H. HANSEN

Mr. Nevin's note contains, I am afraid, a whole bundle of misconceptions. Since, however, the "post-General Theory" literature has evidenced no little confusion with respect to Keynes's interest-rate theory, it may perhaps be worthwhile to comment briefly on matters which by now one might assume had been sufficiently clarified.

Nevin begins with the misconception that I have contended that Keynes convicted the classicists "simply on the grounds that a change in the rate of interest implied a change in the rate of investment." This, of course, is not valid. Acceptance of the investment demand schedule is not an issue here.

Next Nevin asserts that Keynes's point is that any *shift* in the investment demand schedule will cause a shift in the savings schedule (with respect to the rate of interest "to exactly the same degree"). But this is not so. The savings schedule will indeed shift as investment and income rise. The shift in the *schedule* need not, however, be to the same degree. What is true is that the new level of *actual* investment will be matched by an equivalent volume of *actual* saving. But this does not mean that the *schedules* have shifted to the same degree or that the "rate of interest was completely unaffected." Indeed, a shift in the investment demand schedule would in the usual case alter the rate of interest. Productivity *is* a factor in interest rate determination.

9. Keynes, *op. cit.*, chap. 13, p. 172.

It is difficult to make any sense out of the paragraph beginning "Now clearly the Keynesian theory . . . " It seems to be based on a failure to distinguish between observable points in a schedule and the schedule itself (composed of innumerable "virtual" points). Indeed Nevin's note suffers throughout from confusion with respect to this vitally important distinction.

Nevin is quite wrong when he argues that my footnote suggestion (involving a family of I-schedules) leads "more deeply into the morass than Keynes does." Investment may well be a function both of the rate of interest *and* of income, as Hicks indeed had it. But this does not make the IS curve indeterminate as Nevin assumes. A given family of I-curves together with a given family of S-curves definitively fixes an IS curve. Nevin is quite wrong when he says that a family of I-curves together with a family of S-curves suggests a *family* of IS curves.[1]

Nevin is also wrong when he suggests that we can't know the *slope* of the IS curve until we know the *level* of income. This is again simply a muddle. Both the IS curve and the LM curve are known *before* the equilibrium point in the schedules is ascertained. The two schedules consist of virtual points; the intersection of the two curves gives us the observable or equilibrium point which establishes both the level of income and the rate of interest.

Finally it is not true that the difference between Hicks and Keynes may be stated as follows: Keynes, says Nevin, assumed income as constant, while Hicks assumed the quantity of money as constant; one is as good as the other — you can take your choice. This formulation reveals a misconception of Hicks's analysis and indeed of economic analysis in general. It is one thing to assume the determinants (viz., the basic functions) as given, it is another thing to assume the determinates (income and the rate of interest) as given. The former are observable and verifiable patterns of behavior, which by way of illustration may be assumed to be thus and so in order to demonstrate the character of the analysis. Given the basic Keynesian functions, income and the rate of interest are uniquely determined. But it is not

---

1. A shift in the schedule relating investment to the rate of interest means a shift in the whole family of I-schedules. Similarly a shift in the schedule relating saving to the rate of interest means a shift in the entire family of S-schedules. In like manner, a shift in the schedule relating the demand for money to the rate of interest will cause a shift in the whole family of L-schedules — one schedule for each income level. And we could also, if so desired, add a family of M-schedules (money supply); this would have the effect of making the LM curve more elastic.

permissible to assume that you already know the *answer* before you begin your analysis.

The difference between Keynes and Hicks is *not* that Hicks holds the quantity of money constant while Keynes holds income constant. In Hicks's analysis shifts may occur in any one of four determinants of i and Y — i.e., (a) in I-schedules, (b) in S-schedules, (c) in L-schedules, and (d) in M supply. *Changes* in any of these factors will change i and Y. But Keynes pretends that changes in the I-schedules and in the S-schedules (productivity and thrift) do not affect the rate of interest. *That* is the difference.

In Hicks's analysis the determinants are not rabbits pulled out of a hat. The determinants are behavior patterns (the investment, saving and liquidity preference schedules) which can be observed and verified, or, as in the case of the money supply, an operational factor subject to the control of the monetary authority. Keynes's analysis is, however, a clear case of circular reasoning. First he assumes half of the problem already solved, namely, the level of income. If income is given, then we can solve the interest rate problem by his liquidity preference analysis. .And now forsooth, having found the rate of interest we are then able to ascertain what the level of income will be! This surely is nonsense — exactly as Keynes described the classical theory.

Lerner has, of course, let the cat out of the bag. Seeing at long last that productivity and thrift do after all have something to do with the rate of interest, he introduced his sophisticated "liquidity preference curve" which surreptitiously serves as a front for the I-schedules and the S-schedules (see p. 149 in my *Guide to Keynes*).

Once again, however, I wish to point out that Keynes supplied for the first time all the materials needed for an adequate theory. He got himself off in a corner, so brilliantly illumined by his liquidity preference analysis that he lost sight of his own comprehensive system of simultaneous equations — all of which were necessary for a solution of the problem of income and interest rate determination.

ALVIN H. HANSEN.

HARVARD UNIVERSITY

THE MARGINAL AND AVERAGE PROPENSITIES
TO CONSUME: COMMENT

Mr. Hubbard's objective[1] is to explain the apparent contradiction between the observed long-run constancy of the average savings ratio — maintained over a period of growing income *per* head — and the presence of a by no means negligible positive constant in the linear consumption functions obtained from family budget studies and from data of cyclical variations of income and consumption. His explanation of this paradox rests on the alleged property of increments-of-income-due-to-productivity-improvements to induce increments of expenditure larger than other increments of income are capable of doing. True, this could only show that the tendency of a declining average savings ratio, which one is led to expect from family budget and short-run data, would be meeting a counteracting force; it could not show that this force would make the ratio constant. However, once he has got this far, Mr. Hubbard finds no difficulty in supplementing his argument to account for the constancy itself.

Since increased productivity is the cause of long-run income growth, the aptness of Mr. Hubbard's explanation is obvious, should its logic be well founded. But this it is not.

At a given level of employment, a rise of productivity, which is fully reflected in increased money wages (prices remaining constant in consequence), will lead to additional consumption expenditure amounting to some fraction of the wage increment. Mr. Hubbard asserts that this primary increment of expenditure will lead to further expenditure *via* some kind of multiplier process; and since the ratio of the sum of these *two* expenditure increments to the wage increment is obviously higher than whatever the true marginal propensity to consume (as normally understood) is, he believes to have established the special property peculiar to increments of income from higher productivity, which was mentioned at the beginning.

But his reasoning is clearly fallacious. If for the sake of argument we grant that, as Mr. Hubbard suggests, the primary increment of expenditure, induced by higher wages, leads to further expenditures,

1. J. C. Hubbard, "The Marginal and Average Propensities to Consume," this *Journal*, LXVIII (Feb. 1954), 83–96. The substantive part, to which this comment relates, is section II, pp. 85–91.

this it could not do unless additional incomes were created in the first instance. These incomes would then have to be added to the wage increment in the denominator of Mr. Hubbard's ratio, whereupon the alleged exceptionally high value of it would disappear. More fundamentally, however, there will be no additional expenditure at all (beyond the primary increment), because there will be no increase of employment — on the contrary employment will decline — and hence no additional incomes.

Employment will decline because the primary increment of expenditure, induced by the wage increment, will not be sufficient to absorb the additional output forthcoming on account of the increase in productivity — the very increase that was postulated as the cause of the wage increment. There will be, therefore, at the initial level of employment, excess supply. With falling unemployment the primary increment of expenditure will not be maintained, let alone added to by further expenditures. (Mr. Hubbard has a good deal to say on employment changes, but one step too late in the argument.)

In Mr. Hubbard's diagram the initial error — the error which has made possible the building up of an illusory proof of his textual argument — is that the increase of consumption induced by the wage increment is represented as occurring at the level of income ruling prior to the wage increment.

<div align="right">JOHN SPRAOS.</div>

UNIVERSITY OF SHEFFIELD
ENGLAND

FURTHER COMMENT

I have been asked by the Editor to comment on Mr. Spraos' note, and on the issue between him and Mr. Hubbard.

Let us assume there is equilibrium at a constant level of employment and a constant rate of net investment. The increase in the stock of capital will eventually lead to an increase in real output; I shall follow Hubbard in assuming that this increase in real output is fully reflected in higher factor incomes (with constant commodity prices). It seems to me beyond dispute that unless this increase in factor incomes is accompanied by an equal increase in consumption (in Hubbard's words, unless the marginal propensity to consume through a rise in real income is equal to unity), there will be involuntary inventory accumulation resulting in unemployment. Nevertheless a modification of Hubbard's analysis can supply an explanation for the long-run constancy of the consumption-income ratio.

As long as the marginal propensity to consume through increases in real income exceeds the marginal propensity to consume through employment (more briefly, as long as the long-run exceeds the short-run marginal propensity to consume), the short-run consumption function will shift upward with increases in productivity. Such an upward shift is shown in the accompanying figure, where C and I are the consumption and investment functions and OY is equilibrium income. An increase in income from OY to OY'' owing to a rise in productivity will, by hypothesis, bring about an increase in consumption of FB (rather than DB). However, from now on, short-run changes in income will take place only via changes in employment
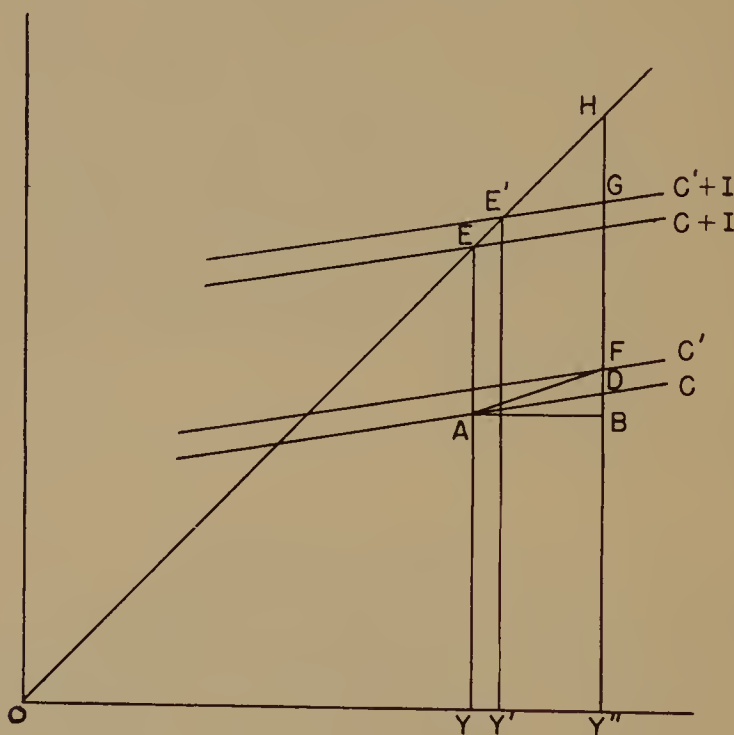
and by hypothesis the marginal propensity to consume through employment remains at DB/AB. This is shown by an upward shift in the short-run consumption function to C', and a movement of the equilibrium point from E to E'. Consequently income is higher than before (OY' rather than OY) but still below the level OY'' required to maintain the initial level of employment. Employment cannot be maintained at OY'' because there will be inventory accumulation to the amount of GH.[1]

1. A similar diagrammatic argument by Peter Gutmann has been brought to my attention.

In the extreme case in which the marginal propensity to consume through a rise in real income is equal to unity (*AF* parallel to *OH*), the initial level of employment will be maintained, and income will be higher at *OY″*.[2] At the other extreme, in which the marginal propensity to consume through real income is equal to the marginal propensity to consume through employment (*FB* = *DB*), the short-run consumption function does not shift and the initial level of income at *OY* will be maintained, but employment will be lower. If, as Professor Hansen has pointed out to me, Kuznets is correct in assuming that the long-run average propensity to consume is constant, we have an intermediate case: the short-run consumption function will shift upward with increases in productivity, but not sufficiently to maintain full employment without an ever-growing rate of investment. So modified, Hubbard's logic can explain such an intermediate case; whether this *particular* intermediate case would result is, of course, an empirical matter.

<div align="right">J. S. CHIPMAN.</div>

UNIVERSITY OF MINNESOTA

REPLY

I believe Mr. Spraos is correct in showing where my analysis in the original article was defective and hence misleading. Professor Hansen's suggestion and Mr. Chipman's interpretation of it as a special case seem to give the proper meaning and focus to what I wanted to say. It seems clear that the marginal propensity to consume through a rise in real income operates as a result of the appearance of new goods, the Duesenberry hypothesis, and other present explanations. Therefore, the revision of spending habits occurs along the margin as real income rises and does not require a complete revision of budget patterns.

<div align="right">J. C. HUBBARD.</div>

BRYN MAWR COLLEGE

2. This takes the place of Hubbard's condition which followed (as Mr. Spraos shows at the end of his note) from a slip in his diagram.

RECENT PUBLICATIONS

ABBOTT, LAWRENCE. Quality and Competition: An Essay in Economic Theory. New York: Columbia University Press. 1955. pp. 229. $3.75.

ALBRECHT, GERHARD. Sozial Politik. Göttingen, Germany: Vandenhoeck and Ruprecht. 1955. pp. 242.

ALTON, THAD PAUL. Polish Postwar Economy. New York: Columbia University Press. 1955. pp. 330. $5.75.

ANGRISANI, GIOVANNI. La Teoria del Commercio internazionale e i Buoni d'Importazione. Rome: Società Tipografica "Leonardo da Vinci." 1955. pp. 316.

BALDWIN, GEORGE B. Beyond Nationalization: The Labor Problems of British Coal. Cambridge, Massachusetts: Harvard University Press. 1955. pp. 324. $6.00.

BENSON, LEE. Merchants, Farmers, and Railroads: Railroad Regulation and New York Politics, 1850–1887. Cambridge, Massachusetts: Harvard University Press. 1955. pp. 310. $5.50.

BURNS, ARTHUR ROBERT. Comparative Economic Organization. New York: Prentice-Hall, Inc. 1955. pp. 766. $7.50.

CALCATERRA, ERCOLE. L'Agricoltura nello Sviluppo delle Economie Arretrate. Milan, Italy: Dott. A. Giuffre. 1955. pp. 153.

CARLI, GUIDO. Verso il Multilateralismo degli Scambi e la Convertibilità delle Monete. Rome: Associazione Bancaria Italiana. 1955. pp. 305.

DEANE, PHYLLIS (EDITOR). Bibliography on Income and Wealth. Cambridge, England: Bowes and Bowes. 1955. pp. 80. [Volume V.]

DIESSLIN, HOWARD G. Agricultural Equipment Financing. New York: National Bureau of Economic Research, Inc. 1955. pp. 95. $1.25. [Occasional Paper 50.]

DI NARDI, GIUSEPPE. Economia dello Scambio. Naples, Italy: Casa Editrice Dott. Eugenio Jovene. 1955. pp. 326.

DJOURABTCHI, HASSAN. La Structure Economique de l'Iran. Geneva, Switzerland: Librairie E. Droz. 1955. pp. 223.

EINAUDI, MARIO; BYÉ, MAURICE; AND ROSSI, ERNESTO. Nationalization in France and Italy. Ithaca, New York: Cornell University Press. 1955. pp. 260. $3.50.

FISHER, WALDO E. AND JAMES, CHARLES M. Minimum Price Fixing in the Bituminous Coal Industry. Princeton, New Jersey: Princeton University Press. 1955. pp. 523. $10.00.

FLUMIANI, C. M. An Introduction to the Theory of the Social Insanities. Albuquerque, New Mexico: Institute for Political and Economic Studies. 1955. pp. 40. $2.00.

GORTER, WYTZE. United States Merchant Marine Policies: Some International Economic Implications. Princeton, New Jersey: International Finance Section, Princeton University. 1955. pp. 16. [Essays in International Finance No. 23.]

GRANGER, GILLES-GASTON. Méthodologie Economique. Paris: Presses Universitaires de France. 1955. pp. 422.

GRAYSON, HENRY. The Crisis of the Middle Class. New York: Rinehart and Company, Inc. 1955. pp. 172. $2.75.

KLEIN, L. R. AND GOLDBERGER, A. S. An Econometric Model of the United States, 1929–1952. Amsterdam, The Netherlands: North-Holland Publishing Company. 1955. pp. 165. $4.50.

Krooss, Herman E. American Economic Development. New York: Prentice-Hall, Inc. 1955. pp. 536. $6.00.

Leftwich, Richard H. The Price System and Resource Allocation. New York: Rinehart and Company, Inc. 1955. pp. 372. $4.50.

Paton, William A. and Paton, William A., Jr. Corporation Accounts and Statements: An Advanced Course. New York: The Macmillan Company. 1955. pp. 740. $6.75.

Peach, W. Nelson and Krause, Walter. Basic Data of the American Economy. Homewood, Illinois: Richard D. Irwin, Inc. 1955. pp. 135. $2.00. [Fourth Edition.]

Perry, J. Harvey. Taxes, Tariffs, and Subsidies. Toronto, Canada: University of Toronto Press. 1955. pp. 1087. $25.00. [Two Volumes.]

Powelson, John P. Economic Accounting. New York: McGraw-Hill Book Company, Inc. 1955. pp. 500. $7.50.

Prais, S. J. and Houthakker, H. S. The Analysis of Family Budgets. New York: Cambridge University Press. 1955. pp. 372. $9.00. [University of Cambridge, Department of Applied Economics, Monograph 4.]

Province of Saskatchewan, Royal Commission on Agriculture and Rural Life. Mechanization and Farm Costs. Regina, Saskatchewan. Royal Commission on Agriculture and Rural Life. 1955. pp. 175. [Report No. 2.]

——. The Scope and Character of the Investigation. Regina, Saskatchewan: Royal Commission on Agriculture and Rural Life. 1955. pp. 122. [Report No. 1.]

Rao, V. K. R. V. An Economic Review of Refugee Rehabilitation in India: A Study of Faridabad Township. Delhi, India: University Press, Delhi School of Economics. 1955. pp. 76. [Monograph No. 3.] ·

——. An Economic Review of Refugee Rehabilitation in India: A Study of Kingsway Camp. Delhi, India: University Press, Delhi School of Economics. 1955. pp. 70. [Monograph No. 5.]

——. An Economic Review of Refugee Rehabilitation in India: A Study of Nilokheri Township. Delhi, India: University Press, Delhi School of Economics. 1954. pp. 56. [Monograph No. 1.]

——. An Economic Review of Refugee Rehabilitation in India: A Study of Rajpura Township. Delhi, India: University Press, Delhi School of Economics. 1955. pp. 78. [Monograph No. 6.]

——. An Economic Review of Refugee Rehabilitation in India: A Study of Tripuri Township. Delhi, India: University Press, Delhi School of Economics. 1955. pp. 68. [Monograph No. 7.]

Rieben, Henri. Des Ententes de Maitres de Forces au Plan Schuman. 1954. pp. 556.

Soule, George. Ideas of the Great Economists. New York: Signet, Signet Key and Mentor Books. 1955. pp. 160. 35 cents.

Steinberg, S. H. (Editor). The Statesman's Yearbook, 1955. New York: St. Martin's Press, Inc. 1955. pp. 1620. $8.50.

Stimson, Claude W. and Lazar, Joseph. Recent Cases and Materials in Business Law. Boston, Massachusetts: Houghton Mifflin Company. 1955. pp. 370. $3.50.

Taft, Philip. Economics and Problems of Labor. Harrisburg, Pennsylvania: The Stackpole Company. 1955. pp. 821. $6.50. [Third edition.]

Tax Institute. Financing Metropolitan Government: Symposium. Princeton, New Jersey: Tax Institute, Inc. 1955. pp. 295. $5.00.

United Nations. Glossary of Commodity Terms. New York: Columbia University Press. 1955. pp. 121. $1.25. [U. N. Sales No. 1954.II.F.4.]

——. Economic Developments in the Middle East, 1945 to 1954. New York: Columbia University Press. 1955. pp. 236. $2.50. [U. N. Sales No. 1955.II.C.2.]

UNITED NATIONS. The Quest for Freer Trade. New York: Columbia University Press. 1955. pp. 59. 40 cents. [U. N. Sales No. 1955.II.C.5.]

——. Review of Economic Activity in Africa 1950 to 1954. New York: Columbia University Press. 1955. pp. 146. $1.50. [U. N. Sales No. 1955.II.C.3.]

——. 1955 Survey of Primary Commodity Markets. New York: Columbia University Press. 1955. pp. 122. $1.75. [U. N. Sales No. 1955.II.D.2.]

——. World Economic Report, 1953–54. New York: Columbia University Press. 1955. pp. 163. $1.75. [U. N. Sales No. 1955.II.C.1.]

——. World Weights and Measures. New York: Columbia University Press. 1955. pp. 225. $2.00. [U. N. Sales No. 1955.XVII.2.]

VAN DER VALK, H. M. H. A. The Economic Future of Canada. New York. McGraw-Hill Book Company, Inc. 1955. pp. 206. $4.75.

WALKER, MABEL. Tax Policy: Modern Merchandising and Municipal Finance: Princeton, New Jersey: Tax Institute, Inc. 1955. pp. 31.

WARNER, W. LLOYD AND ABEGGLEN, JAMES C. Occupational Mobility in American Business and Industry, 1928–1952. Minneapolis, Minnesota: University of Minnesota Press. 1955. pp. 315. $5.50.

WERNETTE, PHILIP. The Future of American Prosperity. New York: The Macmillan Company. 1955. pp. 262. $4.50.